

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 08:28:29 ; Search time 34.13 Seconds  
(without alignments)  
27.760 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209  
Sequence: 1 A1SYGNSADAQPYVCAKIGVDAKQINGKNTAVGIYAGYN 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues  
Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	25.4	1235	1 W21572	Rat brain serine e
2	53	25.4	1249	1 W21571	Rat brain homology
3	52	24.9	199	1 R34398	Helicobacter pylori
4	50	23.9	833	1 W27148	HMG-CoA reductase
5	49.5	23.7	255	1 W55201	H. pylori ORF 01cp
6	49.5	23.7	255	1 W55477	H. pylori ORF 07ep
7	49	23.4	540	1 R48667	Chitinase 1. New c
8	48.5	23.2	435	1 R32999	Rat choline kinase
9	48.5	23.2	1156	1 W46856	Bacillus thuringie
10	48	23.0	341	1 R70142	Porcine mutarotase
11	48	23.0	341	1 R72964	Pig kidney cell mu
12	48	23.0	532	1 R47237	Wild-type feline H
13	48	23.0	532	1 R47237	Feline herpesvirus
14	48	23.0	532	1 W60432	Feline herpesvirus
15	47	22.5	541	1 W44165	Entodinium caudatu
16	46.5	22.2	507	1 P50263	Sequence encoded b
17	46	22.0	165	1 W38713	S. pneumoniae glut
18	46	22.0	493	1 R48669	Chitinase derivati
19	46	22.0	778	1 W35004	Thermotoga maritim
20	46	22.0	832	1 W4089	Human HPT-1 protei
21	46	22.0	1287	1 R79944	Helicobacter pylori
22	46	22.0	1291	1 R75201	Tyrosine phosphata
23	46	22.0	1296	1 R41198	CT. Recombinant He
24	46	22.0	2893	1 W1356	Helicobacter polyp
25	46	22.0	2893	1 W98828	Minor ampullate sp
26	45.5	21.8	251	1 R60192	HAPPV p50 spindle
27	45.5	21.8	351	1 R52635	Streptococcus sp. t
28	45.5	21.8	407	1 W62683	Streptococcus pneu
29	45.5	21.8	592	1 W61249	Glucosyltransferase
30	45.5	21.8	1051	1 W52304	HPV-3 JS isolate
31	45.5	21.8	2233	1 W48711	HPV-3 JS isolate
32	45.5	21.8	2233	1 W48712	HPV-3 Vero CP45 v
33	45.5	21.8	2233	1 W48713	HPV-3 Vero CP45 v
34	45	21.5	288	1 W27720	H. pylori VacA pro

35 45 21.5 453 1 W17784 FIV integrase-LexA  
45 21.5 505 1 R63807 Protein disulphide  
36 45 21.5 514 1 W04346 Chamaecyparis obtu  
37 45 21.5 514 1 W42122 Japanese cypress p  
38 45 21.5 542 1 R48668 Chitinase 2. New c  
39 45 21.5 547 1 R67385 Mitochondrial prot  
40 45 21.5 547 1 R67385 Human alanyl-tRNA  
41 45 21.5 668 1 W19668 Renal cancer assoc  
42 45 21.5 1004 1 Y07088 FIV ppr clone 34 O  
43 45 21.5 1124 1 W53668 H. pylori GHPO 374  
44 45 21.5 1290 1 W98269 IgG-Fc binding pro  
45 45 21.5 2594 1 W14748

## ALIGNMENTS

RESULT 1  
W21572  
ID W21572 standard; Protein; 1235 AA.  
AC W21572.  
DT 20-AUG-1997 (first entry)  
DE Rat brain serine ectopeptidase.  
KW Serine ectopeptidase; tripeptidylpeptidase II; TPP II;  
KW alternative splicing; rodent; rat; cerebral cortex;  
KW neuropeptide; cholecystokinin; CCK; inactivation; degradation;  
KW anorexia; schizophrenia; Parkinson's disease; depression;  
KW Irritable bowel syndrome; bulimia; pathological obesity.  
OS Rattus sp.  
FH Key  
FH region 85..1235  
FT Location/Qualifiers  
FT /note="The C-terminal region of the protein encoded  
FT by the alternatively spliced clone has a  
FT sequence identical to this region"  
PD W09635805-A2.  
PN 14-NOV-1996.  
PR 09-MAY-1996; F00700.  
PR 09-MAY-1995; FR-005489.  
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
PI Bancel RB, Bishop PB, Bourgeat P, Chan S, Ganello CR,  
PI Lebland B, Moore AMT, Schwartz JC, Vargas F;  
PI Lihua Z, Rose C;  
PI WPI: 96-518693/51.  
DR N-PSDB: T72654.  
PT Screening medicaments for treating disorders linked to inactivation  
PT of endogenous neuro-peptide(s) - by contacting candidate molecule  
PT with membrane tri:peptidyl-peptidase (homologue) and measuring  
PT enzyme activity  
PS Claim 4: Pages 166-171; 212pp; French.  
CC An enzyme with specificity for cholecystokinin (CCK) substrates  
CC (specifically the non-sulphated CCK8 and the CCK5 peptides) was  
CC purified from rat cerebral cortex membranes using high performance  
CC liquid chromatography. Based on amino acid sequence data from the  
CC purified enzyme, PCR primers were designed to amplify oligonucleotide  
CC probes A and B of lengths 350 and 380 nucleotides, respectively.  
CC Using these probes, two distinct clones were identified in a rat  
CC brain cDNA library. The first clone encoded the protein which is  
CC the rodent homologue of human tripeptidylpeptidase II. The second  
CC clone differed in the 5' region, probably as a result of alternative  
CC splicing, and encodes a protein of the present sequence which appears  
CC to be a serine ectopeptidase. The rat enzymes are preferred for use  
CC in a new method of screening for medicaments for treating disorders  
CC linked to the inactivation or degradation (or being treatable by  
CC retarding physiological degradation) of endogenous neuropeptides. In  
CC the method, a candidate molecule is contacted with a membrane  
CC tripeptidylpeptidase or homologue, and enzyme activity is measured.  
CC Disorders linked to inactivation or degradation of endogenous  
CC neuropeptides, include food intake disorders, cognitive and motor  
CC disorders such as anorexia, schizophrenia, Parkinson's disease and  
CC depression, as well as gastrointestinal transit disorders such as  
CC irritable bowel syndrome, bulimia and pathological obesity.  
SQ Sequence 1235 AA;

Query Match	25.4%;	Score 53;	DB 1;	Length 1235;
Best Local Similarity	36.7%;	Pred. No. 23;		
Matches 11;	Conservative 6;	Mismatches 13;	Indels 0;	Gaps 0;

Oy 10 AQPYVGAKIGVDAKQINGKNTAYGIYAGY 39  
|||:|:|:|:  
Db 1124 AQPHDGAAGDAEAKKEEGESTLESLEY 1153

RESULT	2
ID	W21571 standard; Protein; 1249 AA.
AC	W21571;
DT	20-AUG-1997 (first entry)
DE	Rat brain homologue of human tripeptidylpeptidase II.
KW	Tripeptidylpeptidase II; TPP II; rodent; rat; cerebral cortex; neuropeptide; cholecystokinin; CKK; inactivation; degradation;
KW	anorexia; schizophrenia; Parkinson's disease; depression;
KW	irritable bowel syndrome; bulimia; pathological obesity;
KW	alternative splicing.
OS	<i>Rattus</i> sp.
FN	key
FT	region
FT	99..1249
FT	/note="The C-terminal region of the protein encoded
FT	by the alternatively spliced clone has a
FT	sequence identical to this region"

PN MO9635805-A2.  
 PN 14-NOV-1996.  
 PE 09-MAY-1996: F00700.  
 PR 09-MAY-1995: FR-005489.  
 PA (INRM ) INSERM INST NAT  
 PI Bembal RB, Bishop PB, Bougaet P, Chan S, Ganelin CR:  
 PI Leblond B, Moore ANJ, Schwartz JC, Vargas F;  
 PI Lihua Z, Rose C;  
 PI WPI: 96-518693/51.  
 DR N-PSDB: T72653.  
 DR  
 PT Screening medicaments for treating disorders linked to inactivation  
 PT of endogenous neuro-peptide(s) - by contacting candidate molecule  
 PT with membrane tri:peptidyl-peptidase (homologue) and measuring  
 PT enzyme activity  
 PS Claim 3: Pages 161-165: 212p; French.  
 CC An enzyme with specificity for cholecystokinin (CCK) substrates  
 CC (specifically the non-sulphated CCK8 and the CCK5 peptides) was  
 CC purified from rat cerebral cortex membranes using high performance  
 CC liquid chromatography. Based on amino acid sequence data from the  
 CC purified enzyme, PCR primers were designed to amplify oligonucleotide  
 CC probes A and B of lengths 350 and 380 nucleotides, respectively.  
 CC Using these probes, two distinct clones were identified in a rat  
 CC brain cDNA library. The first clone encoded the protein of the  
 CC present sequence which is the rodent homologue of human tripeptidyl-  
 CC peptidase II. The second clone differed in the 5' region, probably  
 CC as a result of alternative splicing, and appears to be a serine  
 CC ectopeptidase. The rat enzymes are preferred for use in a new  
 CC method of screening for medicaments for treating disorders linked  
 CC to the inactivation or degradation (or being treatable by retarding  
 CC physiological degradation) of endogenous neuropeptides. In the  
 CC method, a candidate molecule is contacted with a membrane tripeptidyl-  
 CC peptidase or homologue, and enzyme activity is measured.  
 CC Disorders linked to inactivation or degradation of endogenous  
 CC neuropeptides, include food intake disorders, cognitive and motor  
 CC disorders such as anorexia, schizophrenia, Parkinson's disease and  
 CC depression, as well as gastrointestinal transit disorders such as  
 CC irritable bowel syndrome, bulimia and pathological obesity.  
 CC Sequence 1249 AA:  
 CC

Query Match	25.48;	Score 53;	DB 1;	Length 1249;
Best Local Similarity	36.78;	Pred. No. 23;		
Matches 11;	Conservative 6;	Mismatches 13;	Indels 0;	Gaps 0;

```
QY      10 AQPYVGAKIGVDAKQINGKNTAYGIYAGY 39
      111: 11 1 :11: 1::1 : 1
Db      1138 AQPHDGAAGDAEAKEEESTLESLETTY 1167
```

RESULT	3
R34398	
ID	R34398 standard; Protein; 199 AA

AC R34398; (first entry)  
DT 03-AUG-1993  
DE Helicobacter pylori ureg gene.  
KW ureB; ureF; ureG; ureH; ureI; stomach; gastric mucosa; gastric ulcer;  
KW duodenal ulcer; acid tolerance; urease operon.  
FH Helicobacter pylori.  
OS  
FH Key Location/Qualifiers  
FT region 8..15  
FT /note= "ATP-binding"  
FN W09307273-A.  
PN 15-APR-1993.  
PD 02-OCT-1992; F00921.  
PE 03-OCT-1991; FR-012198.  
PR (INSP ) INSERM INST NAT SANTE & RECH MED.  
PA (INSP ) INST PASTEUR.  
PI Cussac V, Ferrero R, Labigne A;  
DR WPI: 93-134462/16.  
DR N-PSDB: Q40312.  
PT Helicobacter pylori genes useful in diagnosis, vaccines and  
PT treatment - necessary for the regulation and maturation of urease  
PS Claim 18; Fig 4; 94pp; French.  
CC The ureg gene is one of 5 new urease genes to be identified by  
CC deletion studies in E.coli where all 5 were found to be necessary  
CC for functional urease expression (regulation/maturation). The ureg  
CC gene encodes a protein of estimated mol. wt. 21.7kD. The Ureg  
CC polypeptide of H.pylori has 92% conservation and 59% identity with  
CC the Ureg polypeptide of K.aerogenes. In K.aerogenes the ureg gene  
CC is implicated in the activation of the apoenzyme by incorporation  
CC of nickel in the urease subunits. See also R34395-R34399.  
SQ Sequence 199 AA;

Query Match	24.9%;	Score 52;	DB 1;	Length 199;
Best Local Similarity	40.0%;	Pred. No. 3.4;		
Matches 12;	Conservative 6;	Mismatches 10;	Indels 2;	Gaps 1;

```
QY      6 NSADAQPYVGA--KIGVDAAQINGKNTAY 33
          | | | | | | : : | : | : | : :
Db      145 NKIDLAPYVGADLKVMERDSKKIAPKSPLE 174
```

RESULT	4
ID	W27148
AC	W27148 standard; Protein: 833 AA.
DT	04-DEC-1997 (first entry)
DE	HMG-CoA reductase degradation protein Hrd3p.
KW	3-hydroxy-3-methylglutaryl: coenzyme A, cholesterol; hrd1p; hrd2p; hrd3p; hypercholesterolaemia; yeast.
OS	Saccharomyces cerevisiae.
FH	Key Location/Qualifiers
FT	peptide 1..18
ET	/note= "putative signal sequence"
FT	region 768..789
ET	/note= "Strongly hydrophobic"
PN	WO9707219-A2.
PD	27-FEB-1997.
PF	16-AUG-1996: IBI161.
PR	17-AUG-1995: US-002381.
PA	(REGC ) UNIV CALIFORNIA.
P1	Hampton R, Rine JD;
DR	N-PSSB; T85268.
PT	3-Hydroxy-3-methylglutaryl CoA reductase degradation polypeptide(s)
PT	[useful as therapeutic agents to reduce hypercholesterolaemia
PS	Claim 4; Page 86-87; 132pp; English.
CC	The present sequence represents the 3-hydroxy-3-methylglutaryl (HMG)-CoA reductase degradation (HRD) protein Hrd3p. Hrd proteins can be used to

CC regulate the degradation of HMG-CoA reductase, e.g. as therapeutic  
CC agents to reduce hypercholesterolaemia, and to elucidate how the  
CC cholesterol pathway modulates the degradation of HMG-CoA reductase. In  
CC addition, as a result of their ability to bind the proteasome complex,  
CC antibodies that specifically bind Hrd polypeptides can be used to  
CC isolate the proteasome complex. Further, they can be used in various  
CC assays to identify compounds that modify the degradation of HMG-CoA  
CC reductase independently of the beneficial LDL receptor control axis.  
CC The nucleic acid molecules can be used as molecular probes for the  
CC isolation of homologous nucleic acid molecules and for the detection of  
CC HRD nucleic acid molecules in yeast.  
SQ Sequence 833 AA;

Query Match 23.9%; Score 50; DB 1; Length 833;  
Best Local Similarity 37.8%; Pred. No. 39;  
Matches 14; Conservative 7; Mismatches 10; Indels 6; Gaps 2;

OY 6 NSADAQPYV--GAKIGVDARINKMTAYGIYAGYN 40  
DB 170 DSAKALLTYQRAAQIGNKAQOV---LAKYKSGFN 202

RESULT 5  
W55201  
ID W55201 standard; Protein; 255 AA.  
AC W55201-1998 (first entry)  
DT 15-JUN-1998

DE H. pylori ORF 01cpl141orf2 protein.  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
OS Helicobacter pylori.  
PN W09737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR ) ASTRA AB.

PI Alm RA, Smith D;

PI WPI; 97-503122/46.

DR N-PSDB; V24610.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PS infection and for diagnosis of H. pylori infection

PS Claim 14; Page 453; 1145pp; English.

CC This sequence is a H. pylori protein of unspecified function.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies

CC against the protein can be used in immunoassays to evaluate the abundance

CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or expected

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.  
SQ Sequence 255 AA;

Query Match 23.7%; Score 49.5; DB 1; Length 255;

Best Local Similarity 34.6%; Pred. No. 11;  
Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5;

OY 3 SYG----NSADAQPYVGAKIGVD-AKQIN-----GKNTAYGIY---AGYN 40  
DB 101 TTYGYSYNNANLS-FVGSKLGIMDASQVNNFTYGVGFALNPFESKRGYN 151

RESULT 6  
W55477  
ID W55477 standard; Protein; 255 AA.  
AC W55477;  
DT 30-JUN-1998 (first entry)

DE H. pylori ORF 07ep11916\_5273452\_c3\_31 cell envelope OMP.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;

KW bacterium.

OS Helicobacter pylori.

PN W09737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR ) ASTRA AB.

PI Alm RA, Smith D;

PI WPI; 97-503122/46.

DR N-PSDB; V24866.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PS infection and for diagnosis of H. pylori infection

PS Claims 14,80; Pages 684-685; 1145pp; English.

CC This sequence is a H. pylori cell envelope outer membrane

CC protein (OMP) having a C-terminal tyrosine cluster motif

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The DNA and probes derived from it may be used for the

CC identification of H. pylori in a sample, and the diagnosis of

CC H. pylori infection. Nucleic acid sequences complementary to the

CC DNA act as antisense sequences, and can be used to prevent the

CC translation of H. pylori mRNA. Antibodies against the protein can

CC be used in immunoassays to evaluate the abundance and distribution

CC of H. pylori-specific antigens. The genomic sequence of H. pylori

CC (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were

CC analysed for ORF of at least 180 nucleotides, and the predicted

CC coding regions defined by computer evaluation. To identify likely

CC H. pylori antigens for vaccine development, the amino acid

CC sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having

CC identified and determined the sequences of interest, particular

CC regions can be isolated from H. pylori by PCR amplification for

CC recombinant polypeptide production, e.g. in E. coli hosts.  
SQ Sequence 255 AA;

Query Match 23.7%; Score 49.5; DB 1; Length 255;  
Best Local Similarity 34.6%; Pred. No. 11;  
Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5;

OY 3 SYG----NSADAQPYVGAKIGVD-AKQIN-----GKNTAYGIY---AGYN 40  
DB 101 TTYGYSYNNANLS-FVGSKLGIMDASQVNNFTYGVGFALNPFESKRGYN 151

RESULT 7  
R48667  
ID R48667 standard; Protein; 540 AA.  
AC R48667;

DT 12-OCT-1994 (first entry)  
 DE Chitinase 1.  
 KW Chitinase; derivative; beta-1,4-glycoside bond; chitin; yeast;  
 KM beta-N-acetylhexosaminidase activity.  
 OS Rhizopus oligosporus IF08631.  
 FH Key location/Qualifiers  
 FT peptide 1..22 /note= "signal peptide"  
 FT protein 23..540 /note= "Mature protein"  
 PN J06046849-A.  
 PD 22-FEB-1994.  
 PF 28-JUL-1992; 201427.  
 PR 28-JUL-1992; JP-201427.  
 PA (KAGO ) KAGOME KK.  
 DR WPI; 94-097015/12.  
 DR N-PSDB; 056756.  
 PT New chitinase enzyme and coding sequence - cleaves beta-1,4-glycoside bond of chitin but has no beta-N-acetylhexosaminidase activity  
 PS Claim 1; Page 16-18; 66pp; Japanese.  
 CC The sequences given in R48667-72 represent chitinase derivatives which cleave the beta-1,4-glycoside bond of chitin but have substantially no beta-N-acetylhexosaminidase activity. The DNA sequences encoding these proteins may be introduced into yeast and cultured for the production of the chitinase proteins.  
 CC Sequence 540 AA;  
 SQ

Query Match 23.4%; Score 49; DB 1; Length 540;  
 Best Local Similarity 34.4%; Pred. No. 32;  
 Matches 11; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

OY 5 GNSADAPYVYKIGOVDAKQINKNTAYG 36  
 Db 145 GGSSDRPFQDAVIDGVLDLEGGASGTGNAF 176

RESULT 8  
 ID R32999 standard; Protein; 435 AA.  
 AC R32999;  
 DT 14-JUN-1993 (first entry)  
 DE Rat choline kinase.  
 KW CK; epitope; tumour; diagnostic; antibodies; distribution; liver;  
 KM cranial nerve disease.  
 OS Rattus rattus.  
 FH Key location/Qualifiers  
 FT peptide 139..225  
 FT peptide 278..327  
 FT peptide 328..360  
 FT peptide 398..412  
 PN J05015367-A.  
 PD 26-JAN-1993.  
 PF 17-SEP-1991; 262516.  
 PR 14-SEP-1991; JP-242528.  
 PA (MITK ) MITSUI TOATSU CHEM INC.  
 DR WPI; 93-070176/09.  
 DR N-PSDB; 037687.  
 PT Rat choline kinase having specified DNA sequence - useful for diagnosing cranial nerve diseases, liver diseases, tumours, etc.  
 PS Claim 1; Page 21; 26pp; Japanese.  
 CC The choline kinase (CK) gene may be obt'd. from rat brain, human brain, human liver, chicken liver, egg, soy bean, etc. The rat CK gene is shown, and can be isolated by PCR using oligonucleotides or the amplified DNA fragments, can then be used as probes to detect CK coding sequences. Rat CK may be purified by MMA-agarose affinity chromatography. A large amt. of CK may be produced and antibodies against CK may be obt'd. for use in detecting the distribution of CK in tissues or cells, and the diagnosis of cranial nerve disease,  
 CC liver disease and tumours.  
 CC Sequence 435 AA;  
 SQ

Query Match 23.2%; Score 48.5; DB 1; Length 435;  
 Best Local Similarity 34.4%; Pred. No. 29;  
 Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 1;

OY 2 ISYGNADAPYVYKAK---IGOVDAKQINKN 30  
 Db 20 ISCGSAAPRPGVGOORDAAGELESKOLGGRS 51

RESULT 9  
 ID W46856 standard; Protein; 1156 AA.  
 AC W46856;  
 DT 11-JUN-1998 (first entry)  
 DE Bacillus thuringiensis toxin designated 86B1(a).  
 KW Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;  
 KM Heliothis virescens; Helicoverpa zea.  
 OS Bacillus thuringiensis.  
 PN W09800546-A2.  
 PD 08-JAN-1998.  
 PF 01-JUL-1997; U11658.  
 PR 01-JUL-1996; US-674002.  
 PA (MYCO ) MYCOGEN CORP.  
 PI Marva KE; Schnepf HE; Stockhof BA; Walz M; Wicker C;  
 DR WPI; 98-086971/08.  
 DR N-PSDB; V16515.  
 PT New isolated Bacillus thuringiensis isolate(s) - used to obtain genes encoding toxins which are active against lepidopteran pests  
 PS such as the Black cutworm  
 CC Claim 34; Pages 82-86; 183pp; English.  
 CC The present sequence represents a Bacillus thuringiensis toxin designated 86B1(a) which is active against lepidopteran pests. The toxin isolates can be used for the control of lepidopteran pests such as Agrotis ipsilon (black cutworm), Heliothis virescens and Helicoverpa zea. PCR primers and probes can be derived from the polynucleotide encoding the toxin and used for the amplification and detection of other CC toxin-encoding sequences.  
 CC Sequence 1156 AA;  
 SQ

Query Match 23.2%; Score 48.5; DB 1; Length 1156;  
 Best Local Similarity 25.5%; Pred. No. 98;  
 Matches 14; Conservative 8; Mismatches 16; Indels 17; Gaps 1;

OY 2 ISYGNADAPYVYKAKIGOVDAKQINKNTAYG-----ITYAG 39  
 Db 389 ISCGHTTATOTILGRNIFRVDSQCNLNDITTYGVNRAVFTYHDASEGSRVDEGT 443

RESULT 10  
 ID R70142 standard; Protein; 341 AA.  
 AC R70142;  
 DT 10-NOV-1995 (first entry)  
 DE Porcine mutarotase (MUT) enzyme.  
 KW Porcine mutarotase; MUT; blood glucose levels;  
 KM recombinant production.  
 OS Sus scrofa.  
 PN J07039380-A.  
 PD 10-FEB-1995.  
 PF 30-JUL-1993; 208328.  
 PR 30-JUL-1993; JP-208328.  
 PA (AMANO ) AMANO PHARM KK.  
 DR WPI; 95-117860/16.  
 DR N-PSDB; 083214.  
 PT DNA encoding mutarotase and recombinant constructs comprising this - for the large scale production of mutarotase for its use in the determination of blood glucose levels  
 PS Claim 3; Pages 7-8; 8pp; Japanese.  
 CC 083214 encodes R70142 porcine kidney mutarotase (MUT). The DNA as part of a claimed expression vector can be used for the recombinant prodn. of the MUT enzyme. The MUT enzyme can be used for the



DT 15-FEB-1999 (first entry)  
DE Feline herpesvirus glycoprotein E.  
KW FHV: vaccine: immunisation: feline rhinotracheitis: vector





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 20, 2000, 10:36:35 ; Search time 27.94 Seconds

(without alignments)  
20.707 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209  
Sequence: 1 A1SYGNSADAPRYVGAKIGQYDAKQINCKNTAYGTYAGYN 40Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0  
Maximum DB seq length: 1000000Post-processing: Minimum Match 0%  
Listing first 45 summariesDatabase :  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	24.9	199	2	US-08-211-312-6
2	52	24.9	199	3	US-08-472-285-6
3	48	23.0	532	3	US-08-911-321-11
4	48	23.0	532	3	US-08-911-321-11
5	48	23.0	532	4	PCT-US95-13975-2
6	47.5	22.7	283	2	US-08-852-401-4
7	46	22.0	832	1	US-08-431-560-1
8	46	22.0	832	1	US-08-463-345-1
9	46	22.0	1287	1	US-08-200-232-2
10	46	22.0	1287	4	PCT-US95-02219-2
11	46	22.0	1287	1	PCT-US95-02219A-2
12	45.5	21.8	251	1	US-08-209-747-8
13	45.5	21.8	251	1	US-08-458-298-8
14	45.5	21.8	331	1	US-08-356-180-3
15	45.5	21.8	2233	2	US-08-569-853-1
16	45.5	21.8	2233	2	US-08-569-853-2
17	45	21.5	462	2	US-08-865-597A-2
18	45	21.5	626	2	US-08-956-242-2
19	45	21.5	668	1	US-08-426-236-2
20	44.5	21.3	359	1	US-08-457-997B-2
21	44	21.1	32	2	US-08-462-221-4
22	44	21.1	32	2	US-08-446-692-21
23	44	21.1	32	2	US-08-488-351A-21
24	44	21.1	74	2	US-08-343-443B-11
25	44	21.1	198	2	US-08-943-915-33
26	44	21.1	207	2	US-08-943-915-2
27	44	21.1	207	2	US-08-943-915-5
28	44	21.1	263	2	US-08-752-844-66
29	44	21.1	339	1	US-08-266-451B-22

30	44	21.1	339	2	US-08-748-725-22	Sequence 22, Appl
31	44	21.1	340	2	US-08-355-844-1	Sequence 1, Appl
32	44	21.1	340	4	PCT-US95-16126-1	Sequence 1, Appl
33	43.5	20.8	380	2	US-08-472-659-34	Sequence 34, Appl
34	43.5	20.8	380	2	US-08-474-661-34	Sequence 34, Appl
35	43.5	20.8	380	2	US-08-611-977-34	Sequence 34, Appl
36	43.5	20.8	449	4	PCT-US94-02539-2	Sequence 2, Appl
37	43.5	20.8	449	4	PCT-US94-02539-2	Sequence 2, Appl
38	43.5	20.8	459	2	US-08-673-312-2	Sequence 2, Appl
39	43.5	20.8	471	2	US-08-657-392-31	Sequence 31, Appl
40	43.5	20.8	471	2	PCT-US94-02539-31	Sequence 31, Appl
41	43.5	20.8	547	2	US-08-467-822-35	Sequence 35, Appl
42	43.5	20.8	573	2	US-08-706-209-1	Sequence 1, Appl
43	43.5	20.8	573	4	PCT-US94-06362-1	Sequence 1, Appl
44	43.5	20.8	573	4	PCT-US96-11373-1	Sequence 1, Appl
45	43.5	20.8	573	4	PCT-US96-11375-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-211-312-6  
: Sequence 6, Application US/08211312  
: Patent No. 5986051  
: GENERAL INFORMATION:  
: APPLICANT: LABIGNE, AGNES  
: APPLICANT: CUSSAC, VALERIE  
: APPLICANT: FERRERO, RICHARD  
: TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY  
: TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: P.C.  
: STREET: 1755 S. Jefferson Davis Highway, Suite 400  
: CITY: Arlington  
: STATE: Virginia  
: COUNTRY: U.S.A.  
: ZIP: 22202  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/211.312  
: FILING DATE: 01-JUL-1994  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: FR 91 12198  
: FILING DATE: 03-OCT-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/FR92/00921  
: FILING DATE: 02-OCT-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Oblon, No. 5986051man F.  
: REGISTRATION NUMBER: 24,618  
: REFERENCE/DOCKET NUMBER: 660-075-0XPCT  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703) 413-3000  
: TELEFAX: (703) 413-2220  
: TELEX: 248855 OPAT UR  
: INFORMATION FOR SEQ ID NO: 6:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 199 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-211-312-6  
Query Match 24.9%, Score 52; DB 2; Length 199;

Best Local Similarity 40.0%; Pred. No. 2;  
Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

OY 6 NSADAPYVGA--KIGQVDAKQINGKNTAY 33  
| | | | | | : : | : | : | :  
Db 145 NKIDLAPYVGADLKVMRDSKTIKAAKSPLE 174

## RESULT 2

US-08-472-285-6  
; Sequence 6, Application US/08472285  
; Patent No. 6027878

## GENERAL INFORMATION:

APPLICANT: LABIGNE, AGNES  
APPLICANT: CUSSAC, VALERIE  
APPLICANT: FERRERO, RICHARD  
TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY  
TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,285  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/211,312  
FILING DATE: 01-JUL-1994  
APPLICATION NUMBER: FR 91 12198  
FILING DATE: 03-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR92/00921  
FILING DATE: 02-OCT-1992

## ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NO. 6027878man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-075-0XPCF  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR

## INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 199 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-285-6

Query Match 24.9%; Score 52; DB 3; Length 199;  
Best Local Similarity 40.0%; Pred. No. 2;

Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

OY 6 NSADAPYVGA--KIGQVDAKQINGKNTAY 33  
| | | | | | : : | : | : | :  
Db 145 NKIDLAPYVGADLKVMRDSKTIKAAKSPLE 174

## RESULT 3

US-08-911-321-11  
; Sequence 11, Application US/08911321

; Patent No. 6010703  
; GENERAL INFORMATION:  
; APPLICANT: Roger K. Maes and Stephen J. Spatz  
; TITLE OF INVENTION: Recombinant Poxvirus  
; TITLE OF INVENTION: Vaccine Against  
; TITLE OF INVENTION: Feline Rhinotracheitis  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ian C. McLeod  
; STREET: 2190 Commons Parkway  
; CITY: Okemos  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48864

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911,321  
FILING DATE:  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/096,183  
FILING DATE: July 26, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-166

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: No. 6010703e  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:

## LENGTH: 532

TYPE: Amino Acid  
STRANDEDNESS: Single  
TOPOLOGY: linear

## MOLECULE TYPE:

DESCRIPTION: Polypeptide  
HYPOHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Feline herpesvirus-1  
STRAIN: 1

## INDIVIDUAL ISOLATE: C-27

## CELL TYPE: N/A

## FEATURE:

## NAME/KEY:

## LOCATION:

## IDENTIFICATION METHOD: Deduced Sequence

## OTHER INFORMATION: ge

OY 7 SADAPYVGA--KIGQVDAKQINGKNTAYGTY 36  
| | | | | | : : | : | : | :  
Db 299 SINHPYIEQPANNVDKFLINVTNAGSLT 328

## RESULT 4

US-08-911-321-11  
; Sequence 11, Application US/08911321  
; Patent No. 6010703  
; GENERAL INFORMATION:  
; APPLICANT: Roger K. Maes and Stephen J. Spatz

Query Match 23.0%; Score 48; DB 3; Length 532;  
Best Local Similarity 36.7%; Pred. No. 27;

Matches 11; Conservative 3; Mismatches 16; Indels 0; Gaps 0;



ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Lisa L.  
REGISTRATION NUMBER: 38, 978  
REFERENCE/DOCKET NUMBER: FER2159POO30US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 283 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-852-401-4

Query Match 22.0%; Score 46; DB 1; Length 832;  
Best Local Similarity 36.4%; Pred. No. 95;  
Matches 11; Conservative 5; Mismatches 19; Indels 5; Gaps 1;

OY 6 NSADAOPYV-----GAKIGQVDAKQINGKNTAYGIYAYN 40  
DB 172 NMGSGNPYPTTEVCAGAGSYVIDQALGAGCTYLLINASN 211

RESULT 7  
US-08-431-560-1  
Sequence 1, Application US/08431560  
Patent No. 5620855  
GENERAL INFORMATION:  
APPLICANT: Anne H. Dantzig, et al.  
TITLE OF INVENTION: Mammalian Influx Peptide  
TITLE OF INVENTION: Transporter  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/431,560  
FILING DATE: 01-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/013,462  
FILING DATE: 04-FEB-1993  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-431-560-1

Query Match 22.0%; Score 46; DB 1; Length 832;  
Best Local Similarity 36.4%; Pred. No. 95;  
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 14 VGAKIGQVDAKQINGKNTAYGI 35  
DB 581 IGTKVGNTAKDEGDISYSL 602

RESULT 8

US-08-463-345-1  
Sequence 1, Application US/08463345  
Patent No. 5710018  
GENERAL INFORMATION:  
APPLICANT: Anne H. Dantzig, et al.  
TITLE OF INVENTION: Mammalian Influx Peptide  
TITLE OF INVENTION: Transporter  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,345  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/013,462  
FILING DATE: 04-FEB-1993  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-463-345-1

Query Match 22.0%; Score 46; DB 1; Length 832;  
Best Local Similarity 36.4%; Pred. No. 95;  
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 14 VGAKIGQVDAKQINGKNTAYGI 35  
DB 581 IGTKVGNTAKDEGDISYSL 602

RESULT 9  
US-08-200-232-2  
Sequence 2, Application US/08200232  
Patent No. 5721349  
GENERAL INFORMATION:  
APPLICANT: Cover, Timothy L.  
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
TITLE OF INVENTION: AND RELATED METHODS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200,232  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratl, Gwendolyn D.



REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.023  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1287 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-200-232-2

Query Match 22.0%; Score 46; DB 1; Length 1287;  
Best Local Similarity 40.5%; Pred. No. 1.6e+02;  
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADQPYGAKIGQVDAKINGKNTAYGIYAGYN 40  
Db 1040 YGTSAGVDAYLN--GQVEA--IVGGFGSYG-YSSFN 1070

RESULT 10  
PCT-US95-02219-2  
Sequence 2, Application PC/TUS9502219  
GENERAL INFORMATION:  
APPLICANT: Cover, Timothy L.  
APPLICANT: Blaser, Martin J.  
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
TITLE OF INVENTION: AND RELATED METHODS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02219  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.023  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1287 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-02219-2

Query Match 22.0%; Score 46; DB 4; Length 1287;  
Best Local Similarity 40.5%; Pred. No. 1.6e+02;  
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADQPYGAKIGQVDAKINGKNTAYGIYAGYN 40  
Db 1040 YGTSAGVDAYLN--GQVEA--IVGGFGSYG-YSSFN 1070

RESULT 11  
PCT-US95-02219A-2  
Sequence 2, Application PC/TUS9502219A  
GENERAL INFORMATION:  
APPLICANT: Cover, Timothy L.  
APPLICANT: Tummuru, Murali KR  
APPLICANT: Cao, Ping  
APPLICANT: Thompson, Stuart A.  
APPLICANT: Blaser, Martin J.  
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
TITLE OF INVENTION: AND THE RELATED METHODS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02219A  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200.023  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1287 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-02219A-2

Query Match 22.0%; Score 46; DB 4; Length 1287;  
Best Local Similarity 40.5%; Pred. No. 1.6e+02;  
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADQPYGAKIGQVDAKINGKNTAYGIYAGYN 40  
Db 1040 YGTSAGVDAYLN--GQVEA--IVGGFGSYG-YSSFN 1070

RESULT 12  
US-08-209-747-8  
Sequence 8, Application US/08209747  
Patent No. 5733771  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Colgin, Mark  
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider  
TITLE OF INVENTION: Silk Proteins  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



Tue May 23 16:14:18 2000

us-09-164-714-1.rai

Page 7

```

1  TITLE OF INVENTION:  VACCINES
2  NUMBER OF SEQUENCES:  2
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  Paul A. Stone
5  STREET:  One Metropolitan Square- 16th Floor
6  CITY:  St. Louis
7  STATE:  Missouri
8  COUNTRY:  USA
9
10 ZIP:  63102
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  Floppy disk
14 COMPUTER:  IBM PC compatible
15 OPERATING SYSTEM:  PC-DOS/MS-DOS
16 SOFTWARE:  Patent Release #1.0, Version #1.25
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/569,853
20 FILING DATE:  08-DEC-1995
21 CLASSIFICATION:  424
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Stone, Paul A.
24
25 REGISTRATION NUMBER:  38,628
26 REFERENCE/DOCKET NUMBER:  SLU 4471
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE:  314-231-5400
29 TELEFAX:  314-231-4342
30 INFORMATION FOR SEQ ID NO:  1:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH:  2233 amino acids
33 TYPE:  amino acid
34 STRANDEDNESS:  single
35 TOPOLOGY:  linear
36 MOLECULE TYPE:  protein
37 ORIGINAL SOURCE:
38 ORGANISM:  Human parainfluenza virus 3
39 STRAIN:  JS
40 JS-08-569-853-1

```

		Query Match	21.8%	Score 45.5;	DB 2:	length 2233:
		Best Local Similarity	35.3%	Pred. No.	9;e+02:	
OY	1 AISTGNSAD-AQPIYGARIGCVDAQCINGKNTAY 33	Matches 12; Conservative	6;	Mismatches 15;	Indels 1;	Gaps 1
Dk	438 AISYENADVYSFSGTGFKNRFEPLDLELITY 471					

```
Search completed: May 20, 2000, 12:18:21
Job time: 6106 sec
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 20, 2000, 11:10:21 ; Search time 43.2 Seconds

(without alignments)  
54.287 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209

Sequence: 1 A1SYGNSADAPYVCAKIGQVDAKQINGKNTAYGIYAGYN 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	59.5	28.5	380	T24786	hypothetical prote
2	55.5	26.6	289	S38806	porin opma - Rhodo
3	55.5	26.6	1196	S40908	TYB protein - yeas
4	55.5	26.6	1278	B22671	TYB protein - yeas
5	55.5	26.6	1328	S52481	TYB protein - yeas
6	55.5	26.6	1328	S52894	TYB protein - yeas
7	55.5	26.6	1328	S50948	TYB protein - yeas
8	55.5	26.6	1328	B28097	TYB protein - yeas
9	55.5	26.6	1348	B23496	TYB protein - yeas
10	55.5	26.6	1749	S69972	TYB protein - yeas
11	55.5	26.6	1755	S50641	TYB protein YER138
12	55.5	26.6	1755	S50663	TYB protein - yeas
13	55.5	26.6	1755	S57047	TYB protein - yeas
14	55.5	26.6	1755	S57045	TYB protein - yeas
15	55.5	26.6	1755	S45736	TYB protein - yeas
16	55.5	26.6	1755	S69963	TYB protein - yeas
17	55.5	26.6	1755	S69839	TYB protein - yeas
18	55.5	26.6	1755	S69980	TYB protein - yeas
19	55.5	26.6	1755	S69957	TYB protein - yeas
20	55.5	26.6	1755	S69951	TYB protein - yeas
21	55.5	26.6	1755	S69969	TYB protein - yeas
22	55.5	26.6	1755	S69960	TYB protein - yeas
23	55.5	26.6	1755	S69979	TYB protein - yeas
24	55.5	26.6	1755	S69838	TYB protein - yeas
25	55.5	26.6	1755	S70238	TYB protein - yeas
26	55.5	26.6	1755	S69955	TYB protein - yeas
27	55.5	26.6	1755	S69845	TYB protein - yeas
28	55.5	26.6	1755	S69866	TYB protein - yeas
29	55.5	26.6	1755	S69982	TYB protein - yeas
30	55.5	26.6	1755	S69949	TYB protein - yeas

31	55.5	26.6	1755	2	S61763	TYB protein - yeas
32	55.5	26.6	1755	2	S69975	TYB protein - yeas
33	55.5	26.6	1756	2	S43867	TYB protein - yeas
34	55.5	26.6	1756	2	S69963	TYB protein - yeas
35	55.5	26.6	1770	2	S45842	TYB protein - yeas
36	55.5	26.6	1770	2	S69953	TYB protein - yeas
37	55.5	26.6	1770	2	S69948	TYB protein - yeas
38	55.5	26.6	1770	2	S70233	TYB protein - yeas
39	55.5	26.6	1770	2	S70230	TYB protein - yeas
40	55.5	26.6	1770	2	S69966	TYB protein - yeas
41	55.5	26.6	1770	2	S69950	TYB protein - yeas
42	55.5	26.6	1770	2	S58651	TYB protein - yeas
43	55.5	26.6	1771	2	S53592	TYB protein - yeas
44	55.5	26.6	1793	2	S52601	TYB protein - yeas
45	55.5	26.6	1810	2	S69973	TYB protein - yeas

## ALIGNMENTS

```
RESULT 1
T24786
hypothetical protein ZK1321.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T24786; T27760
R:Gardner, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z19935
A:Accession: T24786
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-380 <WIL>
A:Cross-references: EMBL:Z48717; PIDN:CAA88612.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone T10B9
R:Gardner, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z20415
A:Accession: T27760
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-380 <WT2>
A:Cross-references: EMBL:Z48584; PIDN:CAA88478.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone ZK1321
C:Genetics:
A:Gene: CESP:ZK1321.4
A:Map position: 2
A:Introns: 4/3; 35/2; 85/3; 115/2; 224/2; 328/2; 346/3

Query Match      28.5%; Score 59.5; DB 2; Length 380;
Best Local Similarity 38.6%; Pred. No. 2.7;
Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

QY      3 SYGNSADAPYVCAKIGQVDAKQIN-----GNTAYGIYAG 38
DB      224 SYNNGATSQPY-GAGSGCTPLNQMTFINTSPAPGANGAYGACCG 266

RESULT 2
S38806
porin opma - Rhodopseudomonas blastic
C:Species: Rhodopseudomonas blastic
C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 07-May-1999
C:Accession: S38806; S38793
R:Kreusch, A.; Neubauer, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E.
Protein Sci. 3: 38-63; 1994
A:Title: Structure of the membrane channel porin from Rhodopseudomonas blastic at 2.
A:Reference number: S38806; MUID:94191532
A:Accession: S38806
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-289 <KRE>
```

R.Krausch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E.  
submitted to the Protein Sequence Database, November 1993

A:Reference number: S38793

A:Accession: S38793

A:Molecule type: protein

A:Residues: 1-71,'X',73-82;91-104;109-137;146-163;173-289 <KR>

C:Genetics:

A:Gene: opmA

C:Keywords: homotrimer; membrane protein

Query Match 26.6%; Score 55.5; DB 2; Length 289;  
Best Local Similarity 41.7%; Pred. No. 6.8;

Matches 15; Conservative 3; Mismatches 15; Indels 3; Gaps 1;

Oy 4 YGNSADAPYVGAKIGQV--DAKQINGKNT 39

Db 227 YGNVAFGATTVRAVYSDIDRA--GADTAVGICADY 259

RESULT 3  
S40908  
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.A

N:Alternate names: protein YAR009c

C:Species: Saccharomyces cerevisiae

C:Date: 13-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 22-Oct-1999

C:Accession: S40908

R:Clarke, M.W.; Keng, T.; Storms, R.R.; Zhong, W.; Fortlin, N.; Zeng, B.; Delaney, S.; Qiu

submitted to the EMBL Data Library, November 1993

A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the 4

A:Reference number: S40891

A:Accession: S40908

A:Molecule type: DNA

A:Residues: 1-1196 <CLA>

A:Cross-references: EMBL:L22015; NID:g1339990; PIDN:AAC04967.1; PID:g2564964

C:Genetics:

A:Map position: 1R

A:Mobile element: retrotransposon Ty1.A

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1196;  
Best Local Similarity 43.8%; Pred. No. 28;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKQINGKNT 31

Db 1053 SYGN---QPYKSOIGNITLNGKVIIGKST 1080

RESULT 4  
B2671  
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty912

C:Species: Saccharomyces cerevisiae

C:Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 26-Aug-1999

C:Accession: B2671; S05850

R:Clarke, J.; Farabaugh, P.

Proc. Natl. Acad. Sci. U.S.A. 82, 2829-2833, 1985

A:Title: Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of

A:Reference number: A94039; MUID:85190589

A:Accession: B2671

A:Molecule type: DNA

A:Residues: 1-1378 <CLA>

R:Mellor, J.; Fulton, S.M.; Dobson, M.J.; Wilson, W.; Kingsman, S.M.; Kingsman, A.J.

Nature 313, 243-246, 1985

A:Title: A retrovirus-like strategy for expression of a fusion protein encoded by yeast

A:Reference number: S05890

A:Accession: S05890

A:Molecule type: DNA

A:Residues: 1-6,'I','8-10','K',12-53 <MEU>

A:Cross-references: EMBL:X01736; NID:g4708; PIDN:CAA25874.1; PID:g1326020

C:Genetics:

A:Mobile element: retrotransposon Ty912

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1278;  
Best Local Similarity 43.8%; Pred. No. 30;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKQINGKNT 31

Db 1135 SYGN---QPYKSOIGNITLNGKVIIGKST 1162

RESULT 5  
S52481  
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: protein YML039W

C:Species: Saccharomyces cerevisiae

C:Date: 08-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 05-Nov-1999

C:Accession: S52481

R:Connor, R.; Churcher, C.

submitted to the EMBL Data Library, February 1995

A:Reference number: S52478

A:Accession: S52481

A:Molecule type: DNA

A:Residues: 1-1328 <CON>

A:Cross-references: EMBL:248430; NID:g683664; PIDN:CAA88330.1; PID:e155222; PID:g1326

C:Genetics:

A:Map position: 13L

A:Mobile element: retrotransposon Ty1

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;  
Best Local Similarity 43.8%; Pred. No. 32;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKQINGKNT 31

Db 1185 SYGN---QPYKSOIGNITLNGKVIIGKST 1212

RESULT 6  
S52894  
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-9532

N:Alternate names: protein YM9532.10c; protein YMR045c

C:Species: Saccharomyces cerevisiae

C:Date: 08-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 05-Nov-1999

C:Accession: S52894

R:Odell, C.; Bowman, S.

submitted to the EMBL Data Library, February 1995

A:Reference number: S52885

A:Accession: S52894

A:Molecule type: DNA

A:Residues: 1-1328 <ODE>

A:Cross-references: EMBL:248502; NID:g695715; PIDN:CAA88411.1; PID:e159526; PID:g1326

C:Genetics:

A:Map position: 13R

A:Mobile element: retrotransposon Ty1-9532

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;  
Best Local Similarity 43.8%; Pred. No. 32;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy 3 SYGNSADAPYVGAKIGQV--DAKQINGKNT 31

Db 1185 SYGN---QPYKSOIGNITLNGKVIIGKST 1212

RESULT 7  
S50948  
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: protein YM9827.08; protein YML045W  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 10-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 05-Nov-1999  
 C:Accession: S50948  
 R:Odell, C.; Bowman, S.  
 submitted to the EMBL Data Library, January 1995  
 A:Reference number: S50941  
 A:Accession: S50948  
 A:Molecule type: DNA  
 A:Residues: 1-1328 <ODE>  
 A:Cross-references: EMBL:Z47816; NID:9642303; PIDN:CAA87830.1; PID:e135608; PID:g1326015  
 C:Genetics:  
 A:Map position: 13L  
 A:Mobile element: retrotransposon Ty1  
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;  
 Best Local Similarity 43.8%; Pred. No. 32;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV--DAKQINGKNT 31  
 |||| ||| :|| : ||||  
 Db 1185 SYGN---QPYKSOIGNIYLNGKVIKGKST 1212

RESULT 8

B28097  
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-H3 (Fragment)

C:Species: Saccharomyces cerevisiae  
 C:Date: 03-Nov-1988 #sequence\_revision 25-Apr-1997 #text\_change 25-Apr-1997  
 C:Accession: B28097  
 R:Boeke, J.D.; Elchinger, D.; Castrillon, D.; Fink, G.R.  
 Mol. Cell. Biol. 8, 1432-1442, 1988  
 A:Title: The Saccharomyces cerevisiae genome contains functional and nonfunctional copies  
 A:Reference number: A28097; MUID:88246410  
 A:Accession: B28097  
 A:Molecule type: DNA  
 A:Residues: 1-1328 <BOE>  
 A:Cross-references: EMBL:M18706  
 C:Genetics:  
 A:Mobile element: retrotransposon Ty1-H3  
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1328;  
 Best Local Similarity 43.8%; Pred. No. 32;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV--DAKQINGKNT 31  
 |||| ||| :|| : ||||  
 Db 1185 SYGN---QPYKSOIGNIYLNGKVIKGKST 1212

RESULT 9

B23496  
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-17

C:Species: Saccharomyces cerevisiae  
 C:Date: 20-Aug-1987 #sequence\_revision 20-Aug-1987 #text\_change 25-Apr-1997  
 C:Accession: B23496  
 R:Warmington, J.R.; Waring, R.B.; Newton, C.S.; Indge, K.J.; Oliver, S.G.  
 Nucleic Acids Res. 13, 6679-6693, 1985  
 A:Title: Nucleotide sequence characterization of Ty 1-17, a class II transposon from yeast  
 A:Reference number: A93591; MUID:86041864  
 A:Accession: B23496  
 A:Molecule type: DNA  
 A:Residues: 1-1348 <MAR>  
 A:Cross-references: EMBL:X03840; EMBL:J01333; EMBL:K00633; EMBL:X02991  
 C:Genetics:  
 A:Mobile element: retrotransposon Ty1-17  
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1348;  
 Best Local Similarity 43.8%; Pred. No. 32;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV--DAKQINGKNT 31  
 |||| ||| :|| : ||||  
 Db 1205 SYGN---QPYKSOIGNIFLNGKVIKGKST 1232

RESULT 10

S69972  
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N

N:Alternate names: protein N2453  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 09-Mar-1996 #sequence\_revision 06-Sep-1996 #text\_change 26-Aug-1999  
 C:Accession: S69972  
 R:Berges, P.; Dolignon, F.; Crouzet, M.  
 submitted to the Protein Sequence Database, April 1996  
 A:Reference number: S62975  
 A:Accession: S69972  
 A:Molecule type: DNA  
 A:Residues: 1-1749 <BER>  
 A:Cross-references: EMBL:Z71330; NID:g1301918; PIDN:CAA95924.1; PID:e239892; PID:g130  
 A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu  
 C:Genetics:  
 A:Map position: 14L  
 A:Mobile element: retrotransposon Ty1.N  
 C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1749;  
 Best Local Similarity 43.8%; Pred. No. 42;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV--DAKQINGKNT 31  
 |||| ||| :|| : ||||  
 Db 1606 SYGN---QPYKSOIGNIFLNGKVIKGKST 1633

RESULT 11

S50641  
 TYB protein YER138C - yeast (Saccharomyces cerevisiae) retrotransposon Ty

C:Species: Saccharomyces cerevisiae  
 C:Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 26-Aug-1999  
 C:Accession: S50641  
 R:Dietrich, F.S.  
 submitted to the EMBL Data Library, December 1994  
 A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lamb  
 A:Reference number: S50428  
 A:Accession: S50641  
 A:Molecule type: DNA  
 A:Residues: 1-1755 <DIE>  
 A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB6465.1; PID:g603378; MIPS:YER1  
 C:Genetics:  
 A:Map position: 5R  
 A:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;  
 Best Local Similarity 43.8%; Pred. No. 42;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV--DAKQINGKNT 31  
 |||| ||| :|| : ||||  
 Db 1612 SYGN---QPYKSOIGNIYLNGKVIKGKST 1639

RESULT 12

S50663  
 TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.E

N:Alternate names: protein YER160C  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 26-Aug-1999

C:Accession: S50663; S30812; S53556  
R:Dietch, F.S.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981, and lambda  
A:Reference number: S50428  
A:Accession: S50663  
A:Molecule type: DNA  
A:Residues: 1-1755 <DIE>  
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64687.1; PID:g603400  
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S30812  
A:Accession: S30812  
A:Molecule type: DNA  
A:Residues: 1-230 <MUL>  
A:Cross-references: EMBL:L10718  
C:Genetics:  
A:Map position: 5R  
A:Mobile element: retrotransposon Ty1.E  
C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;  
Best local Similarity 43.8%; Pred. No. 42;  
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGOV---DAKOINGKNT 31

Db 1612 SYGN---QPYKKSQIGNIYLNGKVIYGKST 1639

RESULT 13

S57047 Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR

N:Alternate names: protein J1570; protein YJR029W

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 17-Mar-1999

C:Accession: S57047; S60512

R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herbe

submitted to the Protein Sequence Database, September 1995

A:Reference number: S57040

A:Accession: S57047

A:Molecule type: DNA

A:Residues: 1-1755 <HUA>

A:Cross-references: EMBL:249528

A:Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residue 4

R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herbe

Yeast 11, 1179-1186, 1995

A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading fr

A:Reference number: S60503; MUID:96109930

A:Accession: S60512

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 'TISTTF', 435-1755 <ZAG>

A:Cross-references: EMBL:X87297

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995

A:Note: the difference at the amino end is due to translation from an incorrect reading

C:Genetics:

A:Map position: 10R

A:Mobile element: retrotransposon Ty1.JR

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;  
Best local Similarity 43.8%; Pred. No. 42;  
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGOV---DAKOINGKNT 31

Db 1612 SYGN---QPYKKSQIGNIYLNGKVIYGKST 1639

RESULT 14

“ ”

S57045 Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR

N:Alternate names: protein J1560; protein YJR027W

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 17-Mar-1999

C:Accession: S57045; S60511

R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; He

submitted to the Protein Sequence Database, September 1995

A:Reference number: S57040

A:Accession: S57045

A:Molecule type: DNA

A:Residues: 1-1755 <HUA>

A:Cross-references: EMBL:249526

A:Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residu

R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; He

Yeast 11, 1179-1186, 1995

A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading

A:Reference number: S60503; MUID:96109930

A:Accession: S60511

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 'TISTTF', 435-1755 <ZAG>

A:Cross-references: EMBL:X87297

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995

A:Note: the difference at the amino end is due to translation from an incorrect read1

C:Genetics:

A:Map position: 10R

A:Mobile element: retrotransposon Ty1.JR

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;  
Best local Similarity 43.8%; Pred. No. 42;  
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGOV---DAKOINGKNT 31

Db 1612 SYGN---QPYKKSQIGNIYLNGKVIYGKST 1639

RESULT 15

S45736 Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1

N:Alternate names: protein YBL004w-a; protein YBL0325

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 21-Nov-1997

C:Accession: S45736; S45735

R:Delaveau, T.; Jacq, C.; Perea, J.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45736

A:Accession: S45736

A:Molecule type: DNA

A:Residues: 1-628 <DEL>

A:Cross-references: EMBL:Z35765; MIPS:YBL004w-a

A:Experimental source: strain S288C

R:Rieger, M.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45734

A:Accession: S45735

A:Molecule type: DNA

A:Residues: 617-1755 <RIE>

A:Cross-references: EMBL:Z35765; MIPS:YBL004w-a

A:Experimental source: strain S288C

C:Genetics:

A:Map position: 2L

A:Mobile element: retrotransposon Ty1

C:Superfamily: Tyb protein

Query Match 26.6%; Score 55.5; DB 2; Length 1755;  
Best local Similarity 43.8%; Pred. No. 42;  
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;



Tue May 23 16:14:19 2000

us-09-164-714-1.rpr

Page 5

```

OY      3  SYGNSADQPIYVAGAKIGV---DAKQINGKNT  31
      |||  |||  ::||:  :  |  |  ||:  |
Db     1612 SYGN---QPYKSKQIGNIYLLNGKVIKSGST  1639

```

Search completed: May 20, 2000, 12:19:13  
Job time: 4132 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 11:16:46 ; Search time 30.1 Seconds  
(Without alignments)  
40.472 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209  
Sequence: 1 A1SYGNSADAPYVGAIKIGQVDAKQINGKNTAYGIYAGYN 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_38:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55.5	26.6	289	1	POR1_RHOBL
2	55.5	26.6	1328	1	P39767 rhodopsin
3	55.5	26.6	1328	1	YME4_YEAST
4	55.5	26.6	1328	1	YME4_YEAST
5	55.5	26.6	1328	1	YMT5_YEAST
6	55.5	26.6	1328	1	YMT5_YEAST
7	55.5	26.6	1347	1	YCB9_YEAST
8	55.5	26.6	1755	1	YJZ7_YEAST
9	55.5	26.6	1755	1	YJZ7_YEAST
10	53.5	25.6	493	1	MOO_MYCN
11	51	24.4	199	1	UREG_HELPY
12	50.3	24.2	250	1	YHIO_NERGO
13	50	23.9	121	1	OPA_HAEN
14	50	23.9	325	1	Y05F_BP74
15	49.5	23.7	449	1	RUXG_SCHPO
16	49.5	23.7	449	1	RUXG_SCHPO
17	49.5	23.7	584	1	PHAC_BORPE
18	49.5	23.4	357	1	ALF_SPIO
19	49.5	23.4	358	1	ALF_ARATH
20	49.5	23.4	540	1	CHIL_RHIO
21	49.5	23.4	959	1	SYAC_SCHPO
22	49.5	23.4	1116	1	SLPH_BACBR
23	49.5	23.4	1262	1	TPP2_MOUSE
24	48.5	23.2	453	1	KICH_RAT
25	48.5	23.0	131	1	RS6E_METVA
26	48.5	23.0	204	1	UREG_STRL
27	48.5	23.0	206	1	UREG_STRL
28	48.5	23.0	313	1	BGL2_YEAST
29	48.5	23.0	1088	1	DPOD_SOYBN
30	48.5	23.0	1242	1	RPO2_ASFB7
31	47.5	22.7	1842	1	FAS2_SCHPO
32	47.5	22.7	1742	1	PEB_ESCPE
33	47.5	22.7	629	1	VE1_HPV31
34	47.5	22.7	1133	1	LOM1_YEAST

35	47	22.5	241	1	CAMT_PETCR	P28034 petroselin
36	47	22.5	242	1	CAMT_VITVI	043237 vitis vinif
37	47	22.5	351	1	MSS2_YEAST	P40990 saccharomyc
38	47	22.5	750	1	TKS7_YEAST	P34231 saccharomyc
39	47	22.5	788	1	PABS_YEAST	P37254 saccharomyc
40	46.5	22.2	329	1	SYFA_HAEN	P3819 haemophilus
41	46.5	22.2	456	1	GUX1_COCCA	000328 cocchiobolu
42	46.5	22.2	521	1	RRP1_IBVK	P12723 avian infec
43	46.5	22.2	581	1	MAXS_YEAST	P40884 saccharomyc
44	46.5	22.2	635	1	DNAK_TREPA	083246 treponema p
45	46.5	22.2	659	1	VE1_HPV03	P36719 human papil

## ALIGNMENTS

RESULT	1	ALIGNMENT
POR1_RHOBL	STANDARD;	PRT; 289 AA.
ID	P39767;	
AC	01-FEB-1995 (Rel. 31, Created)	
DT	01-FEB-1995 (Rel. 31, Last sequence update)	
DT	15-JUL-1999 (Rel. 38, Last annotation update)	
DE	PORIN.	
GN	OPWA.	
OS	Rhodopseudomonas blattica.	
OC	Bacteria; Proteobacteria; alpha subdivision: Rhodobacter group;	
CC	Rhodobacter.	
RN	[1]	
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).	
RX	MEDLINE; 94191532.	
RA	Kreusch A., Neubauer A., Schiltz E., Weckesser J., Schulz G.E.;	
RT	"Structure of the membrane channel porin from Rhodopseudomonas	
RL	blattica at 2.0-A resolution."	
RL	Protein Sci. 3:58-63(1994).	
RN	[2]	
RP	X-RAY CRYSTALLOGRAPHY (1.96 ANGSTROMS).	
RX	MEDLINE; 95055730.	
RA	Kreusch A., Schulz G.E.;	
RT	"Refined structure of the porin from Rhodopseudomonas blattica.	
RT	Comparison with the porin from Rhodobacter capsulatus."	
RL	J. Mol. Biol. 243:891-905(1994).	
RN	[3]	
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF MUTANTS.	
RX	MEDLINE; 98348035.	
RA	Schmid B., Maveyraud L., Kromer M., Schulz G.E.;	
RT	"Porin mutants with new channel properties."	
RL	Protein Sci. 7:1603-1611(1998).	
CC	-1- FUNCTION: FORMS CHANNELS THAT ALLOW THE PASSIVE DIFFUSION OF SMALL	
CC	HYDROPHILIC SOLUTES UP TO AN EXCLUSION LIMIT OF ABOUT 0.6 KD.	
CC	-1- SUBUNIT: HOMOTRIMER.	
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.	
DR	PIR; S38806; S38806.	
DR	PDB; 1PRN; 24-OCT-94.	
DR	PDB; 2PRN; 13-JAN-98.	
DR	PDB; 3PRN; 12-AUG-98.	
DR	PDB; 5PRN; 12-AUG-98.	
DR	PDB; 6PRN; 12-AUG-98.	
DR	PDB; 7PRN; 12-AUG-98.	
DR	PDB; 8PRN; 12-AUG-98.	
DR	PDB; 1BH3; 12-AUG-98.	
KW	Outer membrane; Transmembrane; Porin; 3D-structure.	
SO	SEQUENCE 289 AA; 30597 MW; 08252D9803A104AC CRC64;	
QY	4 YGNSADAPYVGAIKIGQVDAKQINGKNTAYGIYAGYN 39	
DB	227 YGNYAFGATYTRAYVSDIDRA---GADTAYGIAGNY 259	

```
RESULT 2
YMD9_YEAST STANDARD; PRT; 1328 AA.
AC 003434;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YML039W OR YMR054.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Connor R., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48430; CAA88330.1; -
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34
FT NE_BIND 1204 1211
FT SEQUENCE 1328 AA; 151036 MW; C7D14E1AA675E93 CRC64;
SQ
Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
Db 1185 SYGN---QPYKRSQIGNIVLLNGKVIYGKST 1212
RESULT 3
YME4_YEAST STANDARD; PRT; 1328 AA.
ID YME4_YEAST
AC 004711;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YML044W OR YMR827.08.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL: Z47816; CAA87830.1; -
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34
FT NE_BIND 1204 1211
FT SEQUENCE 1328 AA; 150945 MW; E9B964CD76C281F CRC64;
SQ
Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
Db 1185 SYGN---QPYKRSQIGNIVLLNGKVIYGKST 1212
RESULT 4
YMT5_YEAST STANDARD; PRT; 1328 AA.
ID YMT5_YEAST
AC 004214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YMR045C OR YMR9532.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48502; CAA88411.1; -
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34
FT NE_BIND 1204 1211
FT SEQUENCE 1328 AA; 151187 MW; AA19E50B62B43F95 CRC64;
SQ
Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
Db 1185 SYGN---QPYKRSQIGNIVLLNGKVIYGKST 1212
RESULT 5
YMD0_YEAST STANDARD; PRT; 1328 AA.
ID YMD0_YEAST
AC 004670;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSDOSON TY1 PROTEIN B.
GN TY1B OR YMR050C OR YMR9796.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z49703; CAA89760.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 150967 MW; DACD7A471697DIDD CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKXSOIGNIFLNGKVGKST 1212

RESULT 6
YCB9_YEAST STANDARD: PRT; 1347 AA.
ID YCB9_YEAST
AC P25384;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSPOSON TY1-17 PROTEIN B.
DE TY1B OR YCL019W OR YCL19W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86205247.
RA Warrington J.R., Anwar R., Newlon C.S., Waring R.B., Davies R.W.,
RA Indge K.J., Oliver S.G.;
RT "A 'hot-spot' for Ty transposition on the left arm of yeast
RT chromosome III";
RL Nucleic Acids Res. 14:3475-3485(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86041864.
RA Warrington J.R., Waring R.B., Newlon C.S., Indge K.J., Oliver S.G.;
RT "Nucleotide sequence characterization of Ty 1-17, a class II
RT transposon from yeast.";
RL Nucleic Acids Res. 13:6679-6693(1985).
RN [3]
RP SEQUENCE FROM N.A.
RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Stever L.I.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X03840; CAA27458.1; -
DR EMBL; X59720; E264443; -
DR PIR; B23496; B23496.
DR PIR; S19345; S19345.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT NP_BIND 1223 1230 ATP (POTENTIAL).
FT NP_BIND 1223 1230
SQ SEQUENCE 1347 AA; 154069 MW; AD3660C5E7B282FF CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1347;
Best Local Similarity 43.8%; Pred. No. 18;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
Db 1204 SYGN---QPYKXSOIGNIFLNGKVGKST 1231

RESULT 7
YJZ7_YEAST STANDARD: PRT; 1755 AA.
ID YJZ7_YEAST
AC P47098; P87194;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON TY1 PROTEIN B.
DE TY1B OR YUR027W OR J1560.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96109930.
RA Zagulski M., Babinska B., Gromadka R., Migdalski A., Rytko J.,
RA Sulicka J., Herbert C.J.;
RT "The sequence of 24.3 kb from chromosome X reveals five complete open
RT reading frames, all of which correspond to new genes, and a tandem
RT insertion of a Ty1 transposon.";
RL Yeast 11:1179-1186(1995).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z49526; CAA89553.1; -
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFM; PF01021; TYA; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 461 461 PROTEASE (BY SIMILARITY).
FT NP_BIND 1631 1638 ATP (POTENTIAL).
FT NP_BIND 1631 1638
SQ SEQUENCE 1755 AA; 198615 MW; 04E248A77FD7596F CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1755;
Best Local Similarity 43.8%; Pred. No. 23;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
Db 1612 SYGN---QPYKXSOIGNIFLNGKVGKST 1639
```

```

RESULT 8
Y1Z9_YEAST
ID Y1Z9_YEAST STANDARD; PRT: 1755 AA.
AC P47100; P87195;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE TRANSPOSON TY1 PROTEIN B.
GN TY1B OR JTR029W OR J1570.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96109930.
RA Zagulski M., Babinaka B., Gromacka R., Migdalski A., Rylka J.,
RA Sulicka J., Herbert C.J.;
RT "The sequence of 24.3 kb from chromosome X reveals five complete open
RT reading frames, all of which correspond to new genes, and a tandem
RT insertion of a Ty1 transposon."
RL Yeast 11:1179-1186(1995).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 023.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49528; CAA89556.1; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFAM; PF01021; TYA; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KM Aspartyl protease; ATP-binding.
FT ACT_SITE 461 461 PROTEASE (BY SIMILARITY).
FT NF_BIND 1631 1638 ATP (POTENTIAL).
SQ SEQUENCE 1755 AA; 198592 MW; C8B125BF2967C990 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1755;
Best Local Similarity 43.8%; Pred. No. 23;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGV---DAKIQNGKNT 31
Db 1612 SYGN---QPYKSQLIGNIYLNGKVIQKST 1639

RESULT 9
DPOD_DROME
ID DPOD_DROME STANDARD; PRT: 1092 AA.
AC P54358;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE DELTA CATALYTIC CHAIN (EC 2.7.7.7).
GN POLD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96125196.
RA Chiang C.S., Lehman I.R.;
RT "Isolation and sequence determination of the cDNA encoding DNA
RT polymerase delta from Drosophila melanogaster."
RL Gene 166:237-242(1995).
CC -1- FUNCTION: POSSESSES TWO ENZYMATIC ACTIVITIES: DNA SYNTHESIS
CC (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT DEGRADATES SINGLE
CC STRANDED DNA IN THE 3' TO 5' DIRECTION. REQUIRED WITH ITS

```

```

CC ACCESSORY PROTEINS (PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND
CC REPLICATION FACTOR C (RFC) OR ACTIVATOR 1) FOR LEADING STRAND
CC SYNTHESIS. ALSO INVOLVED IN COMPLETING OKAZAKI FRAGMENTS INITIATED
CC BY THE DNA POLYMERASE ALPHA/PRIMASE COMPLEX (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE ->
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: HETERODIMER WITH SUBUNTS OF 125 KD AND 50 KD. THE 125 KD
CC SUBUNIT CONTAINS THE POLYMERASE ACTIVE SITE AND MOST LIKELY THE
CC ACTIVE SITE FOR THE 3'-5' EXONUCLEASE ACTIVITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X88928; CAA61369.1; -.
DR EXBASE; FB890012066; DNAPol-delta.
DR PRINTS; PR00106; DNAPOLB.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
DR PFAM; PF00136; DNA_pol_B; 1.
KW Transferrase; DNA-directed DNA polymerase; DNA replication;
KM DNA-binding; Hydrolase; Exonuclease; Zinc-finger; Nuclear protein.
FT DOMAIN 4 19 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN_FING 997 1017 C4-TYPE (POTENTIAL).
FT ZN_FING 1046 1064 C4-TYPE (POTENTIAL).
SQ SEQUENCE 1092 AA; 124879 MW; 976B640BB37DBD CRC64;

Query Match 25.8%; Score 54; DB 1; Length 1092;
Best Local Similarity 40.9%; Pred. No. 22;
Matches 9; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 13 YVGAKIGVDKQINGKNTAYG 34
Db 692 FTGAQVKLPCLEISGVTAIG 713

RESULT 10
MOO_MYCTU
ID MOO_MYCTU STANDARD; PRT: 493 AA.
AC 005807;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PROBABLE MALATE:OxINONE OXIDOREDUCTASE (EC 1.1.99.16) (MALATE
DE DEHYDROGENASE [ACCEPTOR]) (MOO).
GN RV2852C OR MTCY24AL.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekla F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sultson J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).

```

```

CC -1- CATALYTIC ACTIVITY: (S)-MALATE + ACCEPTOR = OXALOACETATE + REDUCED
CC ACCEPTOR.
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: CITRIC ACID CYCLE.
CC -1- SIMILARITY: BELONGS TO THE MOO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 295207; CAB08454.1; -.
DR TOBERCULIST; RV2852C; -.
KW Oxidoreductase; FAD.
SQ SEQUENCE 493 AA; 53595 MW; C3B4C469FAF285F CRC64;

Query Match
Best Local Similarity 25.6%; Score 53.5; DB 1; Length 493;
Matches 14; Conservative 8; Mismatches 13; Indels 3; Gaps 2;

OY 4 YGNSADAPYVGA-KIGOVDAKOINGKN-TAYGIYAGYN 40
Db 287 YGFPAPGAPPLGAL--HLDLRFVNGKSWLVFGPYAGWS 322

RESULT 11
UREG_HELPEY STANDARD; PRT; 199 AA.
AC 009066;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UREASE ACCESSORY PROTEIN UREG.
GN UREG OR HP0068.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85P;
RX MEDLINE; 92210488.
RA Cussac V., Ferrero R.L., Labigne A.;
RT *Expression of Helicobacter pylori urease genes in Escherichia coli
RT grown under nitrogen-limiting conditions.*;
RL J. Bacteriol. 174:2466-2473(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-D., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khairi H.G., Glick A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Melman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT *The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.*;
RL Nature 388:539-547(1997).
CC -1- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.
CC -1- SIMILARITY: BELONGS TO THE UREG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M84338; AAA25025.1; -.
DR EMBL: A24198; CAA01726.1; -.
DR EMBL: AE000528; AAD07131.1; -.
DR PIR: D41834; D41834.
DR TIGR: HP0068; -.
DR PFAM: PF01495; HypB_Ureg; 1.
KW Nickel; ATP-binding.
FT NP_BIND 8 15 ATP (POTENTIAL).
FT CONFLICT 102 102 D -> S (IN REF. 1).
FT CONFLICT 167 176 MRGEKPIFT -> IAKSPLEIFP (IN REF. 1).
SQ SEQUENCE 199 AA; 21955 MW; A23FE448EB6A208B1 CRC64;

Query Match
Best Local Similarity 24.4%; Score 51; DB 1; Length 199;
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;

OY 6 NSADAPYVGA-KIGOVDAKOINGK 29
Db 145 NKIDLAPYVGAADLKVMERDSKMRGE 170

RESULT 12
YHIO_NEIGO STANDARD; PRT; 250 AA.
ID YHIO_NEIGO
AC P72077;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 27.3 KD PROTEIN.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11;
RA Carrick C.S., Pyfe J.A.M., Davies J.K.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: STRONG. TO E. COLI YHIO AND H. INFLUENZAE HI0949.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U65994; AAC82508.1; -.
DR Hypothetical protein.
KW SEQUENCE 250 AA; 27316 MW; FC749DB91985763 CRC64;

Query Match
Best Local Similarity 24.2%; Score 50.5; DB 1; Length 250;
Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

OY 11 QPYVGAKI-GQVDAKOINGKNTAYGIYAGY 39
Db 216 RPRUGEHLAGAPAYQYTKSTRFDVILPY 245

RESULT 13
OPA_HAEIN STANDARD; PRT; 121 AA.
ID OPA_HAEIN
AC P45088;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE OPA PROTEIN.
GN OPA OR H11174.

```

```

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Balt C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shrivley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullervack T.R., Hanna M.C., Nguyen D.T., Saudel D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT *Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE OF 39-121 FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 96134971.
RA Preston A., Maskell D., Johnson A., Moxon E.R.;
RT "Altered lipopolysaccharide characteristic of the 169 phenotype in
RT Haemophilus influenzae results from mutations in a novel gene, lsn.";
RL J. Bacteriol. 178:396-402(1996).
CC -1- SIMILARITY: SOME TO N.GONORRHOEA OPACITY PROTEIN OPA66.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U32797; AAC22827.1; -
DR EMBL; U17295; AAA95981.1; -
DR TIGR; H11174; -
SQ SEQUENCE 121 AA; 13153 MW; 6647D8B471F42223 CRC64;

Query Match 23.9%; Score 50; DB 1; Length 121;
Best Local Similarity 26.7%; Pred. No. 8.7;
Matches 12; Conservative 6; Mismatches 17; Indels 10; Gaps 1;

QY 4 YGNSADAQPYVGAKIGVDKQINGK-----NTANGIYAG 38
Db 36 FDNNSKVQPYVGARVATNQPKYTNRAEQKPKSSSDIKLGYVAG 80

RESULT 14
Y05F_BP74 STANDARD; PRT; 325 AA.
AC P39261; Q96218;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOHETICAL 36.7 KD PROTEIN IN NRDC-MOBI INTERGENIC REGION.
GN Y05F OR NRDC.10.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
RN [1]
RP SEQUENCE FROM N.A.
RA Mzhavia N., Marusich E., Djavakhishvili T., Netzel J., Peterson S.,
RA Awaya M., Eldermiller J., Canada D., Tracy J., Galbreath K.,
RA Paddison P., Anderson B., Stidham T., Blattner F., Kuter E.M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U76612; AAB26972.1; -
KW Hypothetical protein.
SQ SEQUENCE 325 AA; 36687 MW; FD6E3A67A3D403BF CRC64;

Query Match 23.9%; Score 50; DB 1; Length 325;
Best Local Similarity 47.4%; Pred. No. 23;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 18 IGVDAKQINGKNTAYGIY 36
Db 126 IGVIEKYSINGKISGIGLY 144

RESULT 15
RUXG_SCHPO STANDARD; PRT; 77 AA.
ID 074966;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROBABLE SMALL NUCLEAR RIBONUCLEOPROTEIN G (SNRNP-G) (SM PROTEIN G).
GN SPBC84.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Beck A., Reinhardt R., Lyne M., Wood V., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE COMMON SM PROTEIN, IS FOUND IN U1 AND U2
CC SNRNPs (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL023706; CAA19285.1; -
DR PFM: PF01423; Sm; 1.
KW Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing.
SQ SEQUENCE 77 AA; 8604 MW; 6468E36CEFA8139A CRC64;

Query Match 23.7%; Score 49.5; DB 1; Length 77;
Best Local Similarity 32.4%; Pred. No. 6.5;
Matches 11; Conservative 4; Mismatches 10; Indels 9; Gaps 1;

QY 16 AKIGVDK-----QINGKNTAYGIYAGYN 40
Db 2 SKAGAPDLKRYLDKQVYVQNGSRKRYGVLRGID 35

Search completed: May 20, 2000, 12:22:44
Job time: 3958 sec

```



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 20, 2000, 11:15:03 ; Search time 46.98 Seconds  
(without alignments)  
59.033 Million cell updates/sec

Title: US-09-164-714-1

Sequence: 1 A1SYGNSADAPYVGAKIGQVDAKQINGKNTAYGIYAGYN 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SPTRMBL\_12:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.potent:\*
- 12: sp.virus:\*
- 13: sp.vertebrate:\*
- 14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63.5	30.4	407	2	P96774 haemophilus
2	59.5	28.5	380	5	Q9XTP9 caenorhabdi
3	55.6	26.8	346	2	051841 actinobacil
4	55.5	26.6	767	3	012357 saccharomyc
5	55.5	26.6	1155	3	007163 saccharomyc
6	55.5	26.6	1196	3	013527 saccharomyc
7	55.5	26.6	1285	3	003934 saccharomyc
8	55.5	26.6	1328	3	099231 saccharomyc
9	55.5	26.6	1328	3	003855 saccharomyc
10	55.5	26.6	1346	3	004345 saccharomyc
11	55.5	26.6	1346	3	005679 saccharomyc
12	55.5	26.6	1347	3	005369 saccharomyc
13	55.5	26.6	1347	3	003494 saccharomyc
14	55.5	26.6	1749	3	099337 saccharomyc
15	55.5	26.6	1755	3	012088 saccharomyc
16	55.5	26.6	1755	3	012112 saccharomyc
17	55.5	26.6	1755	3	012141 saccharomyc
18	55.5	26.6	1755	3	012269 saccharomyc
19	55.5	26.6	1755	3	012273 saccharomyc
20	55.5	26.6	1755	3	012316 saccharomyc

## ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	407 AA.
P96774	P96774			
AC	P96774;			
DT	01-MAY-1997 (TRENBLREL. 03, Created)			
DT	01-MAY-1997 (TRENBLREL. 03, Last sequence update)			
DT	01-NOV-1999 (TRENBLREL. 12, Last annotation update)			
DE	OMPA2.			
GN	OMPA2.			
OS	Haemophilus ducreyi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-35000;			
RX	MEDLINE; 97197543.			
RA	KLEINER-TAIT J., HILFKE T.J., MACIVER I., SPINOLA S.M., RADOLF J.D.,			
RA	HENSEN E.J.,			
RT	"The major outer membrane protein of Haemophilus ducreyi consists of			
RT	two OmpA homologs."			
RL	J. Bacteriol. 179:1764-1773(1997).			
DR	EMBL; U60646; AAB49274.1; -			
DR	HSSP; P02934; 1BXW.			
DR	PFAM; PF00691; OmpA. 1			
DR	PFAM; PF01389; OmpA_membrane; 1.			
DR	PRINTS; PR01021; OMPADOMAIN.			
SO	SEQUENCE 407 AA; 44702 MW; 4C2BA5D7 CRC32;			
Query Match	30.4%;	Score 63.5;	DB 2;	Length 407;
Best local similarity	34.4%;	Pred. No. 1.2;		
Matches	21; Conservative	3; Mismatches	8; Indels	29; Gaps 3;
QY	8 ADAPQ---YVGAKIG-----QVDAKQINGKNTA-----YGIYAG 38			
Db	17 ATAPQADTFYVGAKAGASFRHGINSQFDNKKYKKNKHTAVAKKKGKLNTRDSVITYGVAG 76			
QY	39 Y 39			
Db	77 Y 77			
RESULT	2			

```

O9XTP9
ID 09XTP9 PRELIMINARY; PRT; 380 AA.
AC 09XTP9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ZK1321.4 PROTEIN.
GN ZK1321.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA GARDNER A.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z48717; CAA88612.1; -.
DR EMBL; Z48584; CAA88478.1; JOINED.
DR EMBL; Z48584; CAA88478.1; -.
DR EMBL; Z48717; CAA88478.1; JOINED.
SQ SEQUENCE 380 AA; 40259 MW; 1D99FAD9 CRC32;

Query Match 28.5%; Score 59.5; DB 5; Length 380;
Best Local Similarity 38.6%; Pred. No. 3.7;
Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

OY 3 SYGNSADAPYVGAIGQVDAKQIN-----GKNTAYGIVAG 38
Db 224 SYNGATSQPY-GAGSGGTPLNQMFTINTSPACNGATYGAQCG 266

RESULT 3
O51841 PRELIMINARY; PRT; 346 AA.
AC 051841;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE OUTER MEMBRANE PROTEIN 34 PRECURSOR.
GN OMP34.
OS Actinobacillus actinomycetemcomitans
OS (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NCTC 9710;
RX MEDLINE; 96084499.
RA WHITE P.A., NAIR S.P., KIM M.J., WILSON M., HENDERSON B.;
RT "Molecular characterization of an outer membrane protein of Actinobacillus actinomycetemcomitans belonging to the Ompa family."
RL Infect. Immun. 66:369-372(1998).
DR EMBL; AF005079; AAC00068.1; -.
DR HSP; P02934; IBXW.

```

```

DR PROSITE; PS01068; OMPA; 1.
DR PFAM; PF00691; OMPA; 1.
DR PFAM; PF01389; OMPA_membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
KW Signal.
FT SIGNAL. 1 21 POTENTIAL.
SQ SEQUENCE 346 AA; 36905 MW; 8AB0E0A3 CRC32;

Query Match 26.8%; Score 56; DB 2; Length 346;
Best Local Similarity 35.4%; Pred. No. 9.5;
Matches 17; Conservative 4; Mismatches 11; Indels 16; Gaps 3;

OY 8 ADAOP-----YVGAIGQVDA-----KQ-----INGKNTAYGIVAGY 39
Db 19 AQAPAPQANTFYAGAKAGMASHHGLNQFKQGVSIINNSEAYCVFEGY 66

RESULT 4
O12357 PRELIMINARY; PRT; 767 AA.
AC 012357;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
DE FRAME-SHIFT IN TVB PROBABLY NOT FUNCTIONAL.
GN TV1B.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA RIEGER M., MUELLER-AUER S., BRUECKNER M., SCHAEFER M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MIPS;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z72824; CAA97037.1; -.
DR EMBL; Z72823; CAA97029.1; -.
SQ SEQUENCE 767 AA; 87951 MW; B97D40F4 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 767;
Best Local Similarity 43.8%; Pred. No. 29;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAIGQV---DAKQINGKNT 31
Db 624 SYGN---QPYKSOIGINIFLNGKVIGGKST 651

RESULT 5
O07163 PRELIMINARY; PRT; 1155 AA.
AC 007163;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE (STRAIN JB84A CONTAINING PLASMID PNN162).
GN TVB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-JB84A;
RX MEDLINE; 88246410.
RA BOERKE J.D., EICHINGER D., CASTRILLON D., FINK G.R.;
RT "The Saccharomyces cerevisiae genome contains functional and nonfunctional copies of transposon Ty1."
RL Mol. Cell. Biol. 8:1432-1442(1988).
DR EMBL; M18706; AAA66938.1; -.

```

SQ SEQUENCE 1155 AA; 131935 MW; 41599191 CRC32;  
 Query Match 26.6%; Score 55.5; DB 3; Length 1155;  
 Best Local Similarity 43.8%; Pred. No. 47;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;  
 Oy 3 SYGNSADAPYVAGKIGOV---DAKOINGKNT 31  
 Db 1012 SYGN---QPYKSOIGNITLNGKVGKST 1039  
 RESULT 6  
 ID 013527 PRELIMINARY; PRT; 1196 AA.  
 AC 013527;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE VAR009CP.  
 GN VAR009C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 NC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 85182712.  
 RA DRABKIN H.J., RAJBHANDARY U.L.;  
 RT "Attempted expression of a human initiator tRNA gene in Saccharomyces  
 RT cerevisiae.";  
 RL J. Biol. Chem. 260:5596-5602(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 86028187.  
 RA HIETER P., PRIDMORE D., HEGEMANN J.H., THOMAS M., DAVIS R.W.,  
 RA PHILIPPSEN P.;  
 RT "Functional selection and analysis of yeast centromeric DNA.";  
 RL Cell 42:913-921(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 89128457.  
 RA SLATER M.R., CRAIG E.A.;  
 RT "The SSA1 and SSA2 genes of the yeast Saccharomyces cerevisiae.";  
 RL Nucleic Acids Res. 17:805-806(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 90291999.  
 RA HEYER W.D., RAO M.R., ERDILE L.F., KELLY T.J., KOLODNER R.D.;  
 RT "An essential Saccharomyces cerevisiae single-stranded DNA binding  
 RT protein is homologous to the large subunit of human RP-A.";  
 RL EMBO J. 9:2321-2329(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 91071607.  
 RA WYTE W., KROPP L.H., LAMB J., CROWLEY J.C., KABACK D.B.;  
 RT "Molecular cloning of chromosome I DNA from Saccharomyces cerevisiae:  
 RT isolation, characterization and regulation of the SP07 sporulation  
 RT gene.";  
 RL Gene 95:65-72(1990).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 91353080.  
 RA SCHWETZER B., PHILIPPSEN P.;  
 RT "CDC15, an essential cell cycle gene in Saccharomyces cerevisiae,  
 RT encodes a protein kinase domain.";  
 RL Yeast 7:265-273(1991).  
 RN [7]

RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 92051323.  
 RA DAVIES C.J., HUTCHISON C.A. III.;  
 RT "A directed DNA sequencing strategy based upon Tr3 transposon  
 RT mutagenesis: application to the ADE1 locus on Saccharomyces cerevisiae  
 RT chromosome I.";  
 RL Nucleic Acids Res. 19:5731-5738(1991).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 93066269.  
 RA LEEFEBRE O., CARLES C., CONESA C., SWANSON R.N., BOUET F., RIVA M.,  
 RA SENTENAC A.;  
 RT "TFC3: gene encoding the B-block binding subunit of the yeast  
 RT transcription factor TFC.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10512-10516(1992).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 9311037.  
 RA HIRAGA K., SUZUKI K., TSUCHIYA E., MIYAKAWA T.;  
 RT "Cloning and characterization of the elongation factor EF-1 beta  
 RT homologue of Saccharomyces cerevisiae. EF-1 beta is essential for  
 RT growth.";  
 RL FEBS Lett. 316:165-169(1993).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 95028152.  
 RA CLARK M.W., KENG T., STORMS R.K., ZHONG W., FORTIN N., ZENG B.,  
 RA DELANEY S., OUELLETTE B.F., BARTON A.B., AND KABACK D.B.;  
 RT "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of  
 RT the 42 kbp SP07-CEN1-CDC15 region.";  
 RL Yeast 10:535-541(1994).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 95249563.  
 RA BUSSEY H., KABACK D.B., ZHONG W., VO D.T., CLARK M.W., FORTIN N.,  
 RA HALL J., OUELLETTE B.F., KENG T., BARTON A.B., SU Y., DAVIES C.K.,  
 RA STORMS R.K.;  
 RT "The nucleotide sequence of chromosome I from Saccharomyces  
 RT cerevisiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 95249563.  
 RA JIA Y., CHERRY J.M.;  
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L22015; AAC04967.1; -  
 SQ SEQUENCE 1196 AA; 136731 MW; BA97ERCF CRC32;  
 Query Match 26.6%; Score 55.5; DB 3; Length 1196;  
 Best Local Similarity 43.8%; Pred. No. 49;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;  
 Oy 3 SYGNSADAPYVAGKIGOV---DAKOINGKNT 31  
 Db 1053 SYGN---QPYKSOIGNITLNGKVGKST 1080  
 RESULT 7  
 ID 003934 PRELIMINARY; PRT; 1285 AA.  
 AC 003934;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL PROTEIN (FRAGMENT).  
 GN TYB.  
 OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 RN Saccharomycetaceae; Saccharomyces.  
 AC [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-AB972;  
 RA MURPHY L., HARRIS D.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z68329; CAA92721.1; -  
 KW PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DE Hypothetical protein; Hydrolase; Aspartyl protease.  
 FT NON\_TER 1285 1285  
 SQ SEQUENCE 1285 AA; 146556 MW; 03F439B7 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1285;  
 Best Local Similarity 43.8%; Pred. No. 54;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31  
 Db 1203 SYGN---QPYKSQLGNIFLNGKVIYGKST 1230

RESULT 8  
 ID 099231 PRELIMINARY; PRT; 1328 AA.  
 AC 099231;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL 151.2 KD PROTEIN.  
 GN TYB.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA OLIVER K., SHORE L., HARRIS D.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z68194; CAA92352.1; -  
 DR EMBL: Z68195; CAA92360.1; -  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hypothetical protein; Hydrolase; Aspartyl protease.  
 SQ SEQUENCE 1328 AA; 151193 MW; 094C7A55 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1328;  
 Best Local Similarity 43.8%; Pred. No. 56;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31  
 Db 1185 SYGN---QPYKSQLGNIFLNGKVIYGKST 1212

RESULT 9  
 003855

ID 003855 PRELIMINARY; PRT; 1328 AA.  
 AC 003855;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE TYB PROTEIN.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA MURPHY L., HARRIS D.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B., RAJANDREAM M.A.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z47746; CAA87673.1; -  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolase; Aspartyl protease.  
 SQ SEQUENCE 1328 AA; 151007 MW; AC60FC55 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1328;  
 Best Local Similarity 43.8%; Pred. No. 56;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31  
 Db 1185 SYGN---QPYKSQLGNIFLNGKVIYGKST 1212

RESULT 10  
 ID 004345 PRELIMINARY; PRT; 1346 AA.  
 AC 004345;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL 153.9 KD PROTEIN.  
 GN TYB.YD9673.05C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA CONNOR R., CHURCHER C.M.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z68196; CAA92372.1; -  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hypothetical protein; Hydrolase; Aspartyl protease.  
 SQ SEQUENCE 1346 AA; 153889 MW; E0D66880 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1346;  
 Best Local Similarity 43.8%; Pred. No. 57;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAQPYVGAKIGV---DAKQINGKNT 31  
 Db 1203 SYGN---QPYKSQLGNIFLNGKVIYGKST 1230

RESULT 11  
 005679

ID 005679 PRELIMINARY; PRT: 1346 AA.  
 AC 005679;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE TY B.  
 GN YBL0822.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S 288C;  
 RX MEDLINE: 96076635.  
 RA OBERMAIER B., GASSENHUBER J., PIRAVANDI E., DOMEXY H.;  
 RT "Sequence analysis of a 78.6 kb segment of the left end of  
 RL Saccharomyces cerevisiae chromosome II.";  
 RL Yeast 11:1103-1112(1995).  
 DR EMBL: X79489; CAA55998.1; -;  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolyase; Aspartyl protease.  
 SQ SEQUENCE 1346 AA; 153931 MW; 68F2AB6B CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1346;  
 Best Local Similarity 43.8%; Pred. No. 57;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;  
 QY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31  
 Db 1203 SYGN---QPYKSOIGNIFLNGKVGKST 1230

RESULT 12  
 005369  
 ID 005369 PRELIMINARY; PRT: 1346 AA.  
 AC 005369;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE TRANSPOSOM TY1-17 154. OKD HYPOTHETICAL PROTEIN.  
 GN TYB.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RX MEDLINE: 95400292.  
 RA MURAKAMI Y., NAITOU M., HAGIMARA H., SHIBATA T., OZAWA M.,  
 RA SASANUMA S., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,  
 RA YAMAZAKI M., TASHIRO H., EKI T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae.";  
 RL Nat. Genet. 10:261-268(1995).  
 DR EMBL: D50617; BA09237.1; -;  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolyase; Aspartyl protease.  
 SQ SEQUENCE 1346 AA; 153890 MW; 1A8EBF4B CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1346;  
 Best Local Similarity 43.8%; Pred. No. 57;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31  
 Db 1203 SYGN---QPYKSOIGNIFLNGKVGKST 1230

RESULT 13  
 003494  
 ID 003494 PRELIMINARY; PRT: 1347 AA.

AC 003494;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE HYPOTHETICAL 153.9 KD PROTEIN.  
 GN TYB.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA OLIVER K., SHORE L., HARRIS D.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z68194; CAA92351.1; -;  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hypothetical protein; Hydrolyase; Aspartyl protease.  
 SQ SEQUENCE 1347 AA; 153944 MW; DEA98B36 CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1347;  
 Best Local Similarity 43.8%; Pred. No. 57;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;  
 QY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31  
 Db 1204 SYGN---QPYKSOIGNIFLNGKVGKST 1231

RESULT 14  
 099337  
 ID 099337 PRELIMINARY; PRT: 1749 AA.  
 AC 099337;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE CHROMOSOME XIV READING FRAME ORF YNL055C.  
 GN TY1B.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BERGEEZ P., DOIGNON F., CROUZET M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MIPS;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z71331; CAA95928.1; -;  
 DR EMBL: Z71330; CAA95924.1; -;  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PFMW: PF01021; TYA; 1.  
 KW Hydrolyase; Aspartyl protease.  
 SQ SEQUENCE 1749 AA; 198196 MW; 140C9F2E CRC32;

Query Match 26.6%; Score 55.5; DB 3; Length 1749;  
 Best Local Similarity 43.8%; Pred. No. 78;  
 Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAOPYGAKIGOV---DAKOINGKNT 31  
 Db 1606 SYGN---QPYKSOIGNIFLNGKVGKST 1633

RESULT 15  
 012088

```

ID 012088      PRELIMINARY;      PRT: 1755 AA.
AC 012088;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE CHROMOSOME XII READING FRAME ORF YLR036C.
GN TY1B.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP PARTIAL SEQUENCE FROM N.A.
RA OBERMAIER B., PIRAVANDI E., RINKE M.;
RN Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-0 FROM N.A.
RA KOETTER P., ROSE M., ENTIAN K.D.;
RN Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RN Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 273208; CAA97563.1; -.
DR EMBL; 273207; CAA97560.1; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFAM; PF01021; TYA; 1.
KW Hydrolase; Aspartyl protease.
SQ SEQUENCE 1755 AA; 198460 MW; 41619FC4 CRC32;

Query Match      26.68; Score 55.5; DB 3; Length 1755;
Best Local Similarity 43.8%; Pred. No. 78;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADAPYVGAKIGV--DAKQINGKNT 31
DB 1612 SYGN---OPYRKSQIGNITLNGKVIIGKST 1639

```

Search completed: May 20, 2000, 12:22:02  
 Job time: 4019 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 11:01:59 ; Search time 1501.5 Seconds  
(without alignments)  
-351.799 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543  
Sequence: 1 atgaacttaataaacact.....gcgcatttgcttttaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl1:  
1: gb\_da1:\*  
2: gb\_da2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pl4:\*  
11: gb\_pl5:\*  
12: gb\_pl6:\*  
13: gb\_pl7:\*  
14: gb\_pl8:\*  
15: gb\_pl9:\*  
16: gb\_pl10:\*  
17: gb\_pl11:\*  
18: gb\_pl12:\*  
19: gb\_pl13:\*  
20: gb\_pl14:\*  
21: gb\_pl15:\*  
22: gb\_pl16:\*  
23: gb\_pl17:\*  
24: gb\_pl18:\*  
25: gb\_pl19:\*  
26: gb\_pl20:\*  
27: gb\_pl21:\*  
28: gb\_pl22:\*  
29: gb\_pl23:\*  
30: gb\_pl24:\*  
31: gb\_pl25:\*  
32: gb\_pl26:\*  
33: gb\_pl27:\*  
34: gb\_pl28:\*  
35: gb\_pl29:\*  
36: gb\_pl30:\*  
37: gb\_pl31:\*  
38: gb\_pl32:\*  
39: gb\_pl33:\*  
40: gb\_pl34:\*  
41: gb\_pl35:\*  
42: gb\_pl36:\*  
43: gb\_pl37:\*  
44: gb\_pl38:\*  
45: gb\_pl39:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pl4:\*  
52: gb\_pl5:\*  
53: gb\_pl6:\*  
54: gb\_pl7:\*  
55: gb\_pl8:\*  
56: gb\_pl9:\*  
57: gb\_pl10:\*  
58: gb\_pl11:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	45	8.3	1164	34	TTTAP1A	X76126 T. thermophil
2	45	8.3	1575	34	TTTAP1N	X76125 T. thermophil
3	44.4	8.2	119118	11	HSRRY7CCL	AL049760 Human DNA
4	43.6	8.0	823	7	CAMODUT1	X77925 C. albicans
5	43.6	8.0	122852	41	AC009770	AC009770 Homo sapi
6	43.6	8.0	189286	40	AC008125	AC008125 Homo sapi
7	43.6	8.0	208437	57	AC018655	AC018655 Homo sapi
8	42.2	7.8	308399	44	AC005140	AC005140 Plasmid
9	41.6	7.7	1302	51	AF090946	AF090946 Homo sapi
10	41.4	7.6	172633	45	AC016048	AC016048 Homo sapi
11	40.8	7.5	1277	7	SCU20616	U20616 Saccharomyc
12	40.2	7.4	10976	2	AE001127	AE001127 Borrelia
13	39.4	7.3	846	5	I28501	I28501 Sequence
14	39.4	7.3	873	2	AF077664	AF077664 Borrelia
15	39.4	7.3	891	5	I28499	I28499 Sequence
16	39.4	7.3	934	1	BBO5PB31	X74809 B. burgdorfe
17	39.4	7.3	934	1	BBO5PB31	X74809 B. burgdorfe
18	39.4	7.3	934	1	BBO5PB31	X74809 B. burgdorfe
19	39.4	7.3	934	1	BBO5PB31	X74809 B. burgdorfe
20	39.4	7.3	1653	1	BOROSPAC	L23137 Borrelia bu
21	39.4	7.3	1653	1	BOROSPAC	L23138 Borrelia bu
22	39.4	7.3	1653	1	BOROSPAC	L23139 Borrelia bu
23	39.4	7.3	1653	1	BOROSPAC	L23140 Borrelia bu
24	39.4	7.3	1653	1	BOROSPAC	L23141 Borrelia bu
25	39.4	7.3	1653	1	BOROSPAC	L23142 Borrelia bu
26	39.4	7.3	1653	1	BOROSPAC	L23143 Borrelia bu
27	39.4	7.3	1653	1	BOROSPAC	L23144 Borrelia bu
28	39.4	7.3	1653	1	BOROSPAC	L23145 Borrelia bu
29	39.4	7.3	1653	1	BOROSPAC	L23146 Borrelia bu
30	39.4	7.3	1653	1	BOROSPAC	L23147 Borrelia bu
31	39.4	7.3	1653	1	BOROSPAC	L23148 Borrelia bu
32	39.4	7.3	1653	1	BOROSPAC	L23149 Borrelia bu
33	39.4	7.3	1653	1	BOROSPAC	L23150 Borrelia bu
34	39.4	7.3	1653	1	BOROSPAC	L23151 Borrelia bu
35	39.4	7.3	1653	1	BOROSPAC	L23152 Borrelia bu
36	39.4	7.3	1653	1	BOROSPAC	L23153 Borrelia bu
37	39.4	7.3	1653	1	BOROSPAC	L23154 Borrelia bu
38	39.4	7.3	1653	1	BOROSPAC	L23155 Borrelia bu
39	39.4	7.3	1653	1	BOROSPAC	L23156 Borrelia bu
40	39.4	7.3	1653	1	BOROSPAC	L23157 Borrelia bu
41	39.4	7.3	1653	1	BOROSPAC	L23158 Borrelia bu
42	39.4	7.3	1653	1	BOROSPAC	L23159 Borrelia bu
43	39.4	7.3	1653	1	BOROSPAC	L23160 Borrelia bu
44	39.4	7.3	1653	1	BOROSPAC	L23161 Borrelia bu
45	39.4	7.3	1653	1	BOROSPAC	L23162 Borrelia bu

## ALIGNMENTS

RESULT	1
LOCUS	TETAPIPA 1164 bp mRNA INV 02-SEP-1994
DEFINITION	T.thermophila TAP1 gene polyA signal with multiple stop codons.
ACCESSION	X76126
VERSION	X76126.1 GI:426480
KEYWORDS	tapi gene.
SOURCE	Tetrahymena thermophila.
ORGANISM	Tetrahymena thermophila Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena. 1 (bases 1 to 1164)
REFERENCE	Pearlman, R.E. Submitted (03-NOV-1993) R.E. Pearlman, York University, Dept of Biology, Farquharson Bldg., 4700 Keele St., Downsview, Ontario M3J 1P3, CANADA 2 (bases 1 to 1164) Heinonen, T.Y. and Pearlman, R.E. A germ line-specific sequence element in an intron in Tetrahymena thermophila J. Biol. Chem. 269 (26), 17428-17433 (1994) 94292495
JOURNAL	
AUTHORS	
TITLE	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
FEATURES	
source	Location/Qualifiers 1..1164 /organism="Tetrahymena thermophila" /macronuclear /db_xref="taxon:5911" /chromosome="1" /dev_stage="macronucleus" /clone_lib="plasmid" /note="rearranged" 103..1020 /gene="TAP1" 103..105 /gene="TAP1" /note="stop codon TAA-Q" 961..963 /gene="TAP1" /note="stop codon TAA-Q" 967..969 /gene="TAP1" /note="stop codon TAA-Q" 997..999 /gene="TAP1" /note="stop codon TAA-Q" 1018..1020 /gene="TAP1" /note="stop codon TGA" 1139..1144 /note="addition"
BASE COUNT	414 a 122 c 317 g 311 t
ORIGIN	
Query Match	8.3% Score 45; DB 34; Length 1164;
Best Local Similarity	50.7%; Pred. No. 0.2; Mismatches 105; Indels 0; Gaps 0;
Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;	
Dy	83 attctcgtatgctaaccattatgttgglyccaaaattggtcaagtacagccaagcaaa 142
Dd	653 ATGGATGATGCTTGCAAGCGGTGATGATGCCGAAGATGTTGATAACAAGAATCCG 712
OY	143 tcaacggtagaagacaacgcgttatgttattatgaacggtataacttgaccaaatlttg 202
Dd	713 AAGATGGATGATGCTTCAAGTCGTGATGATGCCGAAGATGTTGATTAACGAAGATG 772
OY	203 gcgttagaaccccaatttttgttcaagcgccaagaattaatgcaagcgtagtcctg 262
Dd	773 GCGATGATGCCCAAGATGTGATGATGCTGAAGACGCTGATGATGCCGAAGATGGTGA 832
OY	263 taaaaggtcatgtgaagtcctttgtgtgcttatg 295

[illegible]

ORIGIN		FEATURES
Query Match	8.3%; Score 45; DB 34; Length 1575;	Location/Qualifiers
Best Local Similarity	50.7%; Pred. No. 0.2;	source
Matches 108:	Conservative 0; Mismatches 105; Indels 0; Gaps 0;	
OY	83 attctcgtatgctcaaccctatgttggtgcacaaattgltcaagtagagccgaagcaaa 142	misc_feature /note="214 copies 2 mer 99 55 conserved"
Db	1064 ATGGTGATGATCGTGAAGACGCTGATGATGCCGAAGATGGTGATGATACCAAGATCCG 1123	repeat_region /note="1500 /note="15 copies 26 mer 56 conserved" 369. .488 /note="3 copies 40 mer 76 conserved"
OY	143 tcaacggtagaacaacgcgttatgttatacagttataacttggccaagtatttg 202	repeat_region /note="214 copies 2 mer 99 55 conserved"
Db	1124 AAGATGGGTGAATGCTGTAAGATGGTGTATGCCTGGAAGATGGTGATACCAAGATG 1183	repeat_region /note="1500 /note="15 copies 26 mer 56 conserved" 369. .488 /note="3 copies 40 mer 76 conserved"
OY	203 gcgttaaacccgaatttgtgttcagacgcgaagaattaatgacgagggtgatctcg 262	repeat_region /note="6 copies 6 mer gtagt 86 conserved"
Db	1184 GCGATGATGCCGAAGTGTATGATGCTGTGAGACGGTGTATGATGCCGAAGATGGTGATG 1243	repeat_region /note="4 copies 40 mer 70 conserved" 2527. .2598 /note="12 copies 6 mer cccccg 69 conserved"
OY	263 taagaagtcatgtgaagctcttgcgtctatg 295	repeat_region /note="Single clone region"
Db	1244 ATAAGAAGATGCCGAAGATGTATGATGCTG 1276	misc_feature /note="Single clone region"
RESULT 3	HSRYR7CCL1 119118 bp PRI 12-DEC-1999	
LOCUS	Human DNA sequence from clone XX-PHYR7CCL1 on chromosome 22 contains	repeat_region /note="37 copies 2 mer cc 66 conserved"
DEFINITION	an STS, GSSS, genomic marker D22S928, tc and ca repeat polymorphisms and a putative Cpg island, complete sequence.	repeat_region /note="15 copies 2 mer tt 90 conserved"
ACCESSION	AL049760	repeat_region /note="Alusg repeat: matches 1. .307 of consensus"
VERSION	AL049760.26 GI:5777587	repeat_region /note="AluDo/FRAM repeat: matches 152. .276 of consensus"
KEYWORDS	HFG: ca repeat polymorphism; CpG island; D22S928; tc repeat polymorphism.	repeat_region /note="MER94 repeat: matches 47. .107 of consensus"
SOURCE	human.	repeat_region /note="AluDo/FRAM repeat: matches 152. .276 of consensus"
ORGANISM	Homo sapiens	repeat_region /note="MER94 repeat: matches 47. .107 of consensus"
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	repeat_region /note="Alusg repeat: matches 1. .295 of consensus"
AUTHORS	1 (bases 1 to 119118)	repeat_region /note="Alusg repeat: matches 1. .295 of consensus"
JOURNAL	Ramsay,H.	repeat_region /note="MER94 repeat: matches 27. .47 of consensus"
COMMENT	Direct Submission Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Requests: clonerequests@sanger.ac.uk On Aug 26, 1999 this sequence version replaced gi:5763790. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the validation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, SW, SWISSPROT, Tr, TREMBL, WP, WORMPEP, Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep IMPORTANT: This sequence is not the entire insert of clone XX-PHYR7CCL1 It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone XX-PHYR7CCL1 is at 1 in this sequence. The true left end of clone CPA-21762 is at 119019 in this sequence. The true right end of clone RPA-753M9 is at 50597 in this sequence.	repeat_region /note="MER94 repeat: matches 27. .47 of consensus"

repeat_region	10871..11171	/note="AluJo repeat: matches 1..303 of consensus"
repeat_region	11190..11487	/note="AluXs repeat: matches 1..298 of consensus"
repeat_region	11610..11726	/note="MIR repeat: matches 13..134 of consensus"
repeat_region	12322..12636	/note="AluXs repeat: matches 1..312 of consensus"
repeat_region	13016..13133	/note="L2 repeat: matches 2616..2748 of consensus"
repeat_region	13419..13704	/note="143 copies 2 mer gg 59 conserved"
repeat_region	13432..13769	/note="13 copies 26 mer 57 conserved"
repeat_region	13878..14208	/note="MTRC repeat: matches 6..368 of consensus"
repeat_region	14224..14886	/note="L1PA10 repeat: matches 5476..6165 of consensus"
repeat_region	14893..15204	/note="AluSp repeat: matches 1..312 of consensus"
repeat_region	15226..15335	/note="L1P3 repeat: matches 5356..5465 of consensus"
repeat_region	15336..15446	/note="MTRC repeat: matches 355..464 of consensus"
repeat_region	15484..15777	/note="AluSg repeat: matches 1..295 of consensus"
misc_feature	complement(15768..16376)	/note="match: GSS: Em:AQ476535"
repeat_region	16387..16541	/note="MIR repeat: matches 73..226 of consensus"
repeat_region	17071..17122	/note="26 copies 2 mer aa 73 conserved"
repeat_region	17222..17308	/note="L2 repeat: matches 2652..2738 of consensus"
repeat_region	17452..17487	/note="18 copies 2 mer tg 91 conserved"
repeat_region	17599..17908	/note="AluY repeat: matches 1..310 of consensus"
repeat_region	18508..18616	/note="L1PA6 repeat: matches 6035..6143 of consensus"
repeat_region	18668..18965	/note="AluXs repeat: matches 1..298 of consensus"
repeat_region	19069..19100	/note="MTR1-INTERNAL repeat: matches 692..723 of consensus"
repeat_region	19494..19849	/note="MTR1-INTERNAL repeat: matches 927..1296 of consensus"
repeat_region	19869..20149	/note="AluJo repeat: matches 18..288 of consensus"
repeat_region	20218..20385	/note="MTR1B repeat: matches 1..180 of consensus"
repeat_region	20386..20664	/note="AluSp repeat: matches 1..303 of consensus"
repeat_region	20685..20736	/note="MTR1B repeat: matches 180..234 of consensus"
repeat_region	20737..21016	/note="AluSs repeat: matches 1..280 of consensus"
repeat_region	21017..21215	/note="MTR1B repeat: matches 234..390 of consensus"
repeat_region	21217..21347	/note="MIR repeat: matches 12..144 of consensus"
repeat_region	21582..21731	/note="L2 repeat: matches 2511..2657 of consensus"
repeat_region	22386..22415	/note="15 copies 2 mer ac 93 conserved"
repeat_region	23579..23881	/note="AluXs repeat: matches 1..310 of consensus"
misc_feature	complement(23985..24510)	/note="match: GSS: Em:AQ529159"
misc_feature	complement(24086..24604)	/note="match: GSS: Em:AQ569063"
repeat_region	24115..24201	

[illegible]

[illegible]

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Wrensford,G., Yu,W., Zhou,X. and Gibbs,R.	Direct Submission	Unpublished	2 (bases 1 to 122852)	Morley K.C.	Direct Submission	Submitted (01-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 9, 1999 this sequence version replaced g1:5822592.						
* NOTE: This is a 'working draft' sequence. It currently						
* consists of 83 contigs. The true order of the pieces						
* is not known and their order in this sequence record is						
* arbitrary. Gaps between the contigs are represented as						
* runs of N, but the exact sizes of the gaps are unknown.						
* This record will be updated with the finished sequence						
* as soon as it is available and the accession number will						
* be preserved.						
1	3284:	contig of 3284 bp in length				
*	3285	3309: gap of unknown length				
*	3310	6343: contig of 3034 bp in length				
*	6344	6368: gap of unknown length				
*	6369	9396: contig of 3028 bp in length				
*	9397	9421: gap of unknown length				
*	9422	12446: contig of 3025 bp in length				
*	12447	12471: gap of unknown length				
*	12472	15221: contig of 2750 bp in length				
*	15222	15246: gap of unknown length				
*	15247	17899: contig of 2653 bp in length				
*	17900	17924: gap of unknown length				
*	17925	20312: contig of 2388 bp in length				
*	20313	20336: gap of unknown length				
*	20337	22705: contig of 2369 bp in length				
*	22706	22729: gap of unknown length				
*	22730	25095: contig of 2366 bp in length				
*	25096	25119: gap of unknown length				
*	25120	27456: contig of 2337 bp in length				
*	27457	27480: gap of unknown length				
*	27481	29761: contig of 2281 bp in length				
*	29762	29785: gap of unknown length				
*	29786	31967: contig of 2182 bp in length				
*	31968	31991: gap of unknown length				
*	31992	34129: contig of 2138 bp in length				
*	34130	34153: gap of unknown length				
*	34154	36287: contig of 2134 bp in length				
*	36288	36311: gap of unknown length				
*	36312	38444: contig of 2133 bp in length				
*	38445	38468: gap of unknown length				
*	38469	40566: contig of 2098 bp in length				
*	40567	40590: gap of unknown length				
*	40591	42605: contig of 2015 bp in length				
*	42606	42629: gap of unknown length				
*	42630	44585: contig of 1956 bp in length				
*	44586	44609: gap of unknown length				
*	44610	46497: contig of 1888 bp in length				
*	46498	46521: gap of unknown length				
*	46522	48376: contig of 1855 bp in length				
*	48377	48400: gap of unknown length				
*	48401	50167: contig of 1767 bp in length				
*	50168	50191: gap of unknown length				
*	50192	51933: contig of 1742 bp in length				
*	51934	51957: gap of unknown length				
*	51958	53692: contig of 1735 bp in length				
*	53693	53716: gap of unknown length				
*	53717	55368: contig of 1652 bp in length				
*	55369	55392: gap of unknown length				
*	55393	57008: contig of 1616 bp in length				
*	57009	57032: gap of unknown length				
*	57033	58616: contig of 1584 bp in length				
*	58617	58640: gap of unknown length				
*	58641	60193: contig of 1553 bp in length				
*	60194	60217: gap of unknown length				
*	60218	61768: contig of 1551 bp in length				

```

* 61769 61792: gap of unknown length
* 61793 63341: contig of 1549 bp in length
* 63342 63365: gap of unknown length
* 63366 64896: contig of 1531 bp in length
* 64897 64920: gap of unknown length
* 64921 66447: contig of 1527 bp in length
* 66448 66471: gap of unknown length
* 66472 67989: contig of 1518 bp in length
* 67990 68013: gap of unknown length
* 68014 69504: contig of 1491 bp in length
* 69505 69528: gap of unknown length
* 69529 70952: contig of 1424 bp in length
* 70953 70976: gap of unknown length
* 70977 72397: contig of 1421 bp in length
* 72398 72421: gap of unknown length
* 72422 73816: contig of 1395 bp in length
* 73817 73840: gap of unknown length
* 73841 75232: contig of 1392 bp in length
* 75233 75256: gap of unknown length
* 75257 76626: contig of 1370 bp in length
* 76627 76650: gap of unknown length
* 76651 78002: contig of 1352 bp in length
* 78003 78026: gap of unknown length
* 78027 79362: contig of 1336 bp in length
* 79363 79386: gap of unknown length
* 79387 80682: contig of 1296 bp in length
* 80683 80706: gap of unknown length
* 80707 81988: contig of 1282 bp in length
* 81989 82012: gap of unknown length
* 82013 83377: contig of 1265 bp in length
* 83378 83301: gap of unknown length
* 83302 84565: contig of 1264 bp in length
* 84566 84589: gap of unknown length
* 84590 85843: contig of 1254 bp in length
* 85844 85867: gap of unknown length
* 85868 87102: contig of 1235 bp in length
* 87103 87126: gap of unknown length
* 87127 88345: contig of 1219 bp in length
* 88346 88369: gap of unknown length
* 88370 89564: contig of 1195 bp in length
* 89565 89588: gap of unknown length
* 89589 90775: contig of 1187 bp in length
* 90776 90799: gap of unknown length
* 90800 91972: contig of 1173 bp in length
* 91973 91996: gap of unknown length
* 91997 93153: contig of 1157 bp in length
* 93154 93177: gap of unknown length
* 93178 94329: contig of 1152 bp in length
* 94330 94353: gap of unknown length
* 94354 95488: contig of 1135 bp in length
* 95489 95512: gap of unknown length
* 95513 96635: contig of 1123 bp in length
* 96636 96659: gap of unknown length
* 96660 97780: contig of 1121 bp in length
* 97781 97804: gap of unknown length
* 97805 98916: contig of 1112 bp in length
* 98917 98940: gap of unknown length
* 98941 100000: contig of 1060 bp in length
* 100001 100024: gap of unknown length
* 100025 100750: contig of 726 bp in length
* 100751 100774: gap of unknown length
* 100775 101811: contig of 1037 bp in length
* 101812 101835: gap of unknown length
* 101836 102872: contig of 1037 bp in length
* 102873 102896: gap of unknown length
* 102897 103914: contig of 1018 bp in length
* 103915 103938: gap of unknown length
* 103939 104940: contig of 1002 bp in length
* 104941 104964: gap of unknown length
* 104965 105958: contig of 994 bp in length
* 105959 105982: gap of unknown length
* 105983 106971: contig of 989 bp in length
* 106972 106995: gap of unknown length

```

```

* 106996 107965: contig of 970 bp in length
* 107966 107989: gap of unknown length
* 107990 108952: contig of 963 bp in length
* 108953 108976: gap of unknown length
* 108977 109936: contig of 960 bp in length
* 109937 109960: gap of unknown length
* 109961 110919: contig of 959 bp in length
* 110920 110943: gap of unknown length
* 110944 111895: contig of 952 bp in length
* 111896 111919: gap of unknown length
* 111920 112865: contig of 946 bp in length
* 112866 112889: gap of unknown length
* 112890 113803: contig of 914 bp in length
* 113804 113827: gap of unknown length
* 113828 114716: contig of 889 bp in length
* 114717 114740: gap of unknown length
* 114741 115620: contig of 880 bp in length
* 115621 115644: gap of unknown length
* 115645 116501: contig of 857 bp in length
* 116502 116525: gap of unknown length
* 116526 117361: contig of 836 bp in length
* 117362 117385: gap of unknown length
* 117386 118203: contig of 818 bp in length
* 118204 118227: gap of unknown length
* 118228 119039: contig of 812 bp in length
* 119040 119063: gap of unknown length
* 119064 119706: contig of 643 bp in length

```

Query Match 8.0%; Score 43.6; DB 41; Length 122852;

Best Local Similarity 47.7%; Pred. No. 0.46; Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```

Qy 56 gtcctaacgcctgcacacgaatgcaattctctgatactcacaacctatgtgtgcga 115
Db 71731 GTCATGATGCTGATGCTGATGATGATGCTGATGCTGATGCTGATGCTGATG 71790

Qy 116 aaattgctaaagtagacgcgaacaaatcaacaggtgaagaacacgctatggtattatg 175
Db 71791 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 71850

Qy 176 caggtataactctgacaaatttggcgtgaacccaattgttgttcaagccca 235
Db 71851 AAGCTTATGCGACTGACAGTCCGATGATGATGATGATGATGATGATGATGATG 71910

Qy 236 aagaatttaatgcagcgcgcgtccgttaaagtgatgtgaatccttgggtctatg 295
Db 71911 AAGCTGATGACGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 71970

Qy 296 gcacatacgtataacttcaatc 321
Db 71971 GTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 71996

```

#### RESULT 6

AC008125

LOCUS 189286 bp DNA 25-SEP-1999

DEFINITION Homo sapiens 12 BAC RPc11-25E2 (Roswell Park Cancer Institute

AC008125 Human BAC library) complete sequence.

AC008125.9 GI:5923640

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 189286)  
Muzny,D.M., Adams,C., Bailey,M., Barabara,J., Blankenburg,K.,  
Bodola,B., Bouck,J., Bowie,S., Brooks,A., Bunney,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,  
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,

Holloway, C., Hosak, H., Issar, A., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kneitz, S., Kondejowski, N., Kong, Y., Kovar, C., Lau, S., Leal, B., Lee, E., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R., Maronde, I., Martin, R., Martinez, C., Meleod, M.P., Mei, G., Merscher, S., Miller, A., Montgomery, K.T., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswald, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A., Stamps, A., Sugang, R., Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahbh, M., Watlington, S., Weinstein, G., Weinstein, K.T., Williamson, A., Worley, K., Wren, J., Wrenford, G., Xiang, A.M., Yang, R., Yu, W., Zhou, X., Kuchelapatti, R., Nelson, D. and Gibbs, R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 189286)  
AUTHORS Worley, K.C.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 189286)  
Worley, K.C.  
TITLE Direct Submission  
JOURNAL Submitted (25-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 25, 1999 this sequence version replaced g1:5851701.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 clones with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig Length: 189286  
Phrap values in estimate: 188412  
Average error rate (BCM-Phrap estimate): 0.000392177  
Fraction of Phrap values less than 40 : 0.0664395

Number of consensus changing edits:  
Number of N's in consensus : 22  
0

Position	Original+Context	Edited+Context
5887	atcacacct(n)caagttccct	atcacacct(c)caagttccct
42337	aatgcttct(n)caagtcag	aatgcttct(c)caagtcag
47267	ttttttaga(n)ggagcttgc	ttttttaga(t)ggagcttgc
55095	tgatgaagc(n)gtttccatc	tgatgaagc(c)gtttccatc
74098	caagcttct(n)nnntccctt	caagcttct(g)nnntccctt
74099	agcttctct(n)nnntccctt	agcttctct(g)nnntccctt
74100	gcttctctm(n)ntcccttca	gcttctctg(c)ntcccttca
74101	cttctctnn(n)tcctcttca	cttctctgc(g)tcctcttca
77151	tgaagatgg(n)aaagatatc	tgaagatgg(a)aaagatatc
77153	agaagatgga(n)agatatcaa	agaagatgga(a)agatatcaa
95463	ttgaagagcc(n)cccaagttca	ttgaagagcc(c)cccaagttca
103775	gttaacatagg(a)atacagttc	gttaacatagg(c)atacagttc
104001	atgaatgtt(n)tagcttcatc	atgaatgtt(c)tagcttcatc
151027	ccttagaac(n)cacnaattac	ccttagaac(a)cacnaattac
151031	agaaacnac(n)aattactaac	agaaacnac(a)aattactaac
151057	ctcaagagga(n)aacgaatac	ctcaagagga(a)aacgaatac
151154	ccaatggtt(n)aaagaagaat	ccaatggtt(a)aaagaagaat
151643	cccaacagtaa(n)aatccagag	cccaacagtaa(g)aatccagag
173642	tatatatagt(n)tatatatata	tatatatagt(a)tatatatata
173684	atgatctt(n)tgagacacac	atgatctt(c)tgagacacac
179756	ggagccccc(n)ccaaacacgt	ggagccccc(a)ccaaacacgt

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
10001		*	*	*	*	*	*	*	*
9001		*	*	*	*	*	*	*	*
8001		*	*	*	*	*	*	*	*
7001		*	*	*	*	*	*	*	*
6001		*	*	*	*	*	*	*	*
5001		*	*	*	*	*	*	*	*
4001		*	*	*	*	*	*	*	*
3001		*	*	*	*	*	*	*	*
2001		*	*	*	*	*	*	*	*
1001		*	*	*	*	*	*	*	*
01		*	*	*	*	*	*	*	*

Version: 1.01 gxf0.

#### FEATURES Location/Qualifiers

source 1..189286  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RPC111-25E2"  
58..174  
/rpt\_family="MER103"  
complement(309..376)  
/rpt\_family="Alu"  
1586..1653  
/rpt\_family="MIR"  
complement(1693..1817)  
/rpt\_family="MIR"  
1882..2059  
/rpt\_family="L2"  
complement(2681..2813)  
/rpt\_family="L2"  
complement(3681..3898)  
/rpt\_family="MIR"  
4057..4368  
/rpt\_family="AluSp"  
6005..6025  
repeat\_region





```
*
* 6660 6679: gap of unknown length
* 6680 7547: contig of 868 bp in length
* 7548 7567: gap of unknown length
* 7568 8375: contig of 808 bp in length
* 8376 8395: gap of unknown length
* 8396 9213: contig of 818 bp in length
* 9214 9233: gap of unknown length
* 9234 9400: contig of 167 bp in length
* 9401 9421: gap of unknown length
* 9421 50801: contig of 41381 bp in length
* 50802 50821: gap of unknown length
* 50822 95943: contig of 45122 bp in length
* 95944 95963: gap of unknown length
* 95964 208437: contig of 112474 bp in length.
*
FEATURES
  Source
    1. .208437
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="12q"
      /clone="RP11-946G22"
BASE COUNT  61426 a 43297 c 43806 g 59518 t 390 others
ORIGIN
Query Match      8.0%; Score 43.6; DB 57; Length 208437;
Best Local Similarity 47.7%; Pred. No. 0.46;
Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

Qy  56  gtcgaacgctgcacatgcagctatgcgaattcgtcgtatgctcaacccattgtgtgcga 115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  80546 GTGATGATGCGATGCGATGCGATGCGTGTACTGATGATGCGAGATGCGTATGCTGATG 80487

Qy  116  aaattgctcaagtagacgcaagaacaaacggtlaagaacacccgtatggtattatg 175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  80486 ATGATGCTGTATGCTGCTGATGATGATGCTTATGCTGTAATGATGATGATGATGATG 80427

Qy  176  caggttaacttgaccacaaatttggcgtagaacccgaattgtgtgtcgaagcga 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  80426 ATGGTATGCGAGTGCAGTGCATGCTGATGTAATGATGATGTTATGCGCAGTGCATGCTG 80367

Qy  236  aagaatttaatgcagcgctgctcgtlaaagtgatgtgaagctcttctgtctattg 295
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  80366 ATGCTGATGACGATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 80307

Qy  296  gccacatcgcataactcaatcaat 321
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  80306 GTAATGATGATTAATATATGATGATGAT 80281

RESULT  8
AC005140/c 308399 bp DNA HTG 01-JAN-2000
LOCUS Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
ACCESSION AC005140.5 GI:6652490
VERSION AC005140.5
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 308399)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Tamaki, T., Kurd, O.B.,
          Conway, A.B. and Davis, R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
          2 (bases 1 to 308399)
AUTHORS Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology
          Center, Stanford University, 855 California Avenue, Palo Alto, CA
          94304, USA
COMMENT On Jan 1, 2000 this sequence version replaced g1:5919273.
* NOTE: This is a 'working draft' sequence. It currently
```

```
*
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 66875 66874: contig of 66874 bp in length
* 67075 67074: gap of unknown length
* 177925 177925: contig of 110851 bp in length
* 177926 178125: gap of unknown length
* 178126 225300: contig of 47175 bp in length
* 225301 225500: gap of unknown length
* 225501 241505: contig of 16005 bp in length
* 241506 241705: gap of unknown length
* 241706 247313: contig of 5608 bp in length
* 247314 247513: gap of unknown length
* 247513 257513: contig of 10000 bp in length
* 257514 257713: gap of unknown length
* 257714 277104: contig of 19391 bp in length
* 277105 277305: gap of unknown length
* 277305 308399: contig of 31095 bp in length.
*
FEATURES
  Source
    1. .308399
      /organism="Plasmodium falciparum"
      /db_xref="taxon:5833"
      /chromosome="12"
      /clone="3D7"
BASE COUNT  124664 a 29405 c 30157 g 122771 t 1402 others
ORIGIN
Query Match      7.8%; Score 42.2; DB 44; Length 308399;
Best Local Similarity 44.2%; Pred. No. 1;
Matches 173; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

Qy  115  aaattgctcaagtagacgcaagaacaaatcaaggtlaagaacacccgtatggtattat 174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  239260 AAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239201

Qy  175  gcaagtttaacttgaccacaaatttggcgtagaacccgaattgtgtgtcgaagcgc 234
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  239200 GATGATGCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239141

Qy  235  aagaatttaatgcagcgctgagtcctcgtlaaagtgatgtgaagctcttctgtctattg 294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  239140 AATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239081

Qy  295  gccacatcgcataactcaatcaatccccattttagccaaggcaattaggaatt 354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  239080 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 239021

Qy  355  gctaaactaaagtagagtagtccagcgccgttaatgaactaactcaaaaagaagcagc 414
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  239020 GATGACCAATGTGAATAAAGAAACCAATTAATTAATTAATTAATTAATTAATTAAT 238961

Qy  415  aaacccagccagcagcgcgctgtgtgtcgttlaaacattagccaatgtggcgct 474
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  238960 ATTAAGGATTAATCAACAGCTATTCGCAACTTGTAAAGAAAGTAAGAAAGAAATGACGCTA 238901

Qy  475  gaagcaagctacaactatctatcagaagatg 505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  238900 CAAGATGATTAATCAATTAATGAAGAAAGAAAGAAATG 238870

RESULT  9
AF090946/c 1302 bp mRNA PRI 12-JAN-2000
LOCUS AF090946 Homo sapiens clone H00688.
DEFINITION AF090946
ACCESSION AF090946.1 GI:6690254
KEYWORDS FLI_CDNA.
SOURCE human.
```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1302)  
AUTHORS Yu,Y., Zhang,C., Luo,L., Ouyang,S., Zhang,S., Li,W., Wu,J.,  
Zhou,S., Liu,M., and He,F.  
TITLE Functional prediction of the coding sequences of 50 new genes  
deduced by analysis of cDNA clones from human fetal liver  
Unpublished  
JOURNAL 2 (bases 1 to 1302)  
REFERENCE Yu,Y., Zhang,C., Luo,L., Ouyang,S., Zhang,S., Li,W., Wu,J.,  
Zhou,S., Liu,M., and He,F.  
AUTHORS Direct Submission  
TITLE Submitted (10-SEP-1998) Dept. of Experimental Hematology, Beijing  
JOURNAL Institute of Radiation Medicine, 27 Taiping RD, Beijing 100850,  
P.R.China  
FEATURES  
source Location/Qualifiers  
1. 1302  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HQ0688"  
/tissue\_type="liver"  
/dev\_stage="fetus"  
BASE COUNT 324 a 217 c 179 g 415 t 167 others  
ORIGIN  
Query Match 7.7%; Score 41.6; DB 51; Length 1302;  
Best Local Similarity 15.3%; Pred. No. 1.4;  
Matches 38; Conservative 108; Mismatches 102; Indels 0; Gaps 0;  
OY 24 ggcagcatcagctctctgtattgagcagtgagtgctaacgctgcacatcagcatggcaa 83  
DB 793 GACGACAGTGYKGRARRAMWMMKRMMAAKKSYTTKSCMAARWYMWTVWARGSYNG 734  
OY 84 tctgcgtgctcctcacccctatgtgtgcacaaattgtgcagtgatgagccagcaagaat 143  
DB 733 GKTTRKMSMTMYCAMMYRRAMWMTTTTMYKSMRSCCRMRKRAKKSSKSMK 674  
OY 144 caacgtaagaacacgcctatgattatgcaaggtataacttgacaaattgg 203  
DB 673 GRMCWYMGRAARGRMWARGSYCYMRGYSKMRBARARMTYMMMMRRMTTR 614  
OY 204 cgtagaaccgaattgtgtgttcagagccaaagaattaatgcagcgctgagtcctgt 263  
DB 613 AARAGRMWMTWMTWGSYKRAKRAKMAKRAAMWMSRARRMTTWAARRGARAGCTYKR 554  
OY 264 aaaaggtg 271  
DB 553 AAKMSMG 546

RESULT 10  
AC016048/C  
LOCUS AC016048 172633 bp DNA HTG 21-JAN-2000  
DEFINITION Homo sapiens chromosome 17 clone RP11-98G11, WORKING DRAFT  
ACCESSION AC016048  
VERSION AC016048.1 GI:6456151  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 172633)  
AUTHORS Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,  
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Matathe,R.,  
Morehouse,A.U., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,  
Yu,S., and Davis,R.W.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 172633)  
AUTHORS Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,  
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Matathe,R.,

COMMENT  
TITLE  
JOURNAL  
Morehouse,A.U., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,  
Yu,S., and Davis,R.W.  
Direct Submission  
Submitted (20-NOV-1999) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
----- Genome Center  
Center: Stanford DNA Sequencing and Technology Development  
Center  
Center code: SPSSTDC  
Web site: http://sequence-www.stanford.edu/group/human/  
Contact: hum.info@sequence.stanford.edu  
----- Project Information  
Center project name: 705  
Center clone name: RP11-98G11  
----- Summary Statistics  
Sequencing Vector: M13mp18; X02513  
Chemistry: Dye-Primer; % of reads  
Chemistry: Dye-Terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 154547 bases at least Q40  
Consensus quality: 162256 bases at least Q30  
Consensus quality: 164167 bases at least Q20  
Insert size: 177356; agarose-fp  
Insert size: 171633; sum-of-contigs  
Quality coverage: 5.2x in Q20 bases; agarose-fp  
Quality coverage: 5.3x in Q20 bases; sum-of-contigs.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1248: contig of 1248 bp in length  
\* 1249 1298: gap of unknown length  
\* 1299 2671: contig of 1373 bp in length  
\* 2672 2721: gap of unknown length  
\* 2722 4038: contig of 1317 bp in length  
\* 4039 4089: gap of unknown length  
\* 4089 5264: contig of 1175 bp in length  
\* 5264 6770: gap of unknown length  
\* 6770 6820: contig of 1457 bp in length  
\* 6820 9900: gap of unknown length  
\* 9900 9901: contig of 3080 bp in length  
\* 9901 9951: gap of unknown length  
\* 9951 13563: contig of 3613 bp in length  
\* 13564 13613: gap of unknown length  
\* 13614 18849: contig of 5236 bp in length  
\* 18850 18899: gap of unknown length  
\* 18900 24187: contig of 5288 bp in length  
\* 24188 24337: gap of unknown length  
\* 24338 29472: contig of 5235 bp in length  
\* 29473 29522: gap of unknown length  
\* 29523 35549: contig of 6027 bp in length  
\* 35550 35599: gap of unknown length  
\* 35600 42925: contig of 7326 bp in length  
\* 42926 42975: gap of unknown length  
\* 42976 48502: contig of 5527 bp in length  
\* 48503 48552: gap of unknown length  
\* 48553 57258: contig of 8706 bp in length  
\* 57259 57308: gap of unknown length  
\* 57309 66203: contig of 8895 bp in length  
\* 66204 66254: gap of unknown length  
\* 66255 73236: contig of 8963 bp in length  
\* 73237 75286: gap of unknown length  
\* 75287 85275: contig of 9989 bp in length  
\* 85276 85325: gap of unknown length  
\* 85326 95182: contig of 9857 bp in length  
\* 95183 95232: gap of unknown length  
\* 95233 118006: contig of 22774 bp in length  
\* 118007 118056: gap of unknown length

\* 118057 144296: contig of 26240 bp in length  
\* 144297 144346: gap of unknown length  
\* 144347 172633: contig of 28287 bp in length.

FEATURES  
Source  
1. 172633  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/clone="RP11-98611"  
/clone\_1bp="RPI human BAC library 11"

BASE COUNT 40832 a 46710 c 45544 g 38472 t 1075 others  
ORIGIN

Query Match 7.6%; Score 41.4; DB 45; Length 172633;  
Best Local Similarity 48.5%; Pred. No. 1.7;  
Matches 114; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 77 atggcaattcgtcgatgctcaaacccctatgtgtgcccagaattggtcaagtagagcgca 136  
Db 76270 ATGGTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76211  
Qy 137 agcaaatcaacggtgaagaacacccgttatgtattatcagcagttataacttgaccaa 196  
Db 76210 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76151  
Qy 197 atttgcgttagaaccgaattgtgtgttcagacgcgaagaatttaatcagagcgctga 256  
Db 76150 ATGATGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 76091  
Qy 257 gtccctgtaaaaggtgtagtgaagtcctttgtgtgttattagacatatcgcatata 311  
Db 76090 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 76036

## RESULT 11

SCU20616 1277 bp DNA PLN 22-JUL-1995  
LOCUS SCU20616  
DEFINITION Saccharomyces cerevisiae G4p2 gene, complete cds.  
ACCESSION U20616.1 GI:710337  
VERSION U20616.1 GI:710337  
KEYWORDS  
SOURCE baker's yeast.  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycetales;  
Saccharomycetaceae; Saccharomyces.

## REFERENCE

REFERENCE  
AUTHORS Frantz,J.D. and Gilbert,W.  
TITLE A yeast gene product, G4p2, with a specific affinity for quadruplex  
nucleic acids  
JOURNAL J. Biol. Chem. 270 (16), 9413-9419 (1995)  
MEDLINE 95238458  
REFERENCE 2 (bases 1 to 1277)  
AUTHORS Frantz,J.D.  
TITLE Direct Submission  
JOURNAL Submitted (03-FEB-1995) J. Daniel Frantz, Biology, Harvard  
University, Cambridge, MA 02138, USA  
LOCATION/Qualifiers  
1. 1277

## FEATURES

source  
/organism="Saccharomyces cerevisiae"  
/strain="S288c"  
/db\_xref="taxon:4932"  
/clone="pool of clones G4p2-1, -2, -3"  
/clone\_1bp="Yeast genomic DNA in lambda EMBL3a; Snyder et.  
al. PNAS 83, p.730"  
234. 1055  
/note="Similar to yeast suppressor protein MPT4,  
Swiss-Prot Accession Number P39015"  
/product="G4p2"  
/codon\_start=1  
/protein\_id="AA070169.1"  
/db\_xref="GI:710338"  
/translation="MSNPFDLLGNVEDADVDVYLPPKEIYKSNSTSKKADVPPPSADP

SKARNRPRSGNEGATRDKTPAGRRNRNSKDVDTATTKSNFRATDRSRGKTDPT  
KKKXNCGWGDKKRELSAEKQADAAAEADADAGDAPKPAQSLSDYLNQAN  
NOKKXPEAKEVEVLDAGRIETAEAEAYPATKAVNYSKOLKTEIELEFDIVESNT  
RKNGGDDNNRNNRNRNGRGARKGNMTNANANVSANTYOKNRNIDVSNLPSLA"  
BASE COUNT 415 a 262 c 250 g 350 t  
ORIGIN

Query Match 7.5%; Score 40.8; DB 7; Length 1277;  
Best Local Similarity 44.7%; Pred. No. 2.3;  
Matches 159; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 91 gatgctaacccctatgtgtgtgtgcccagaattgtgtcagtagagcgcaagaatcaatcaggt 150  
Db 597 GAAGTCAAGCCGATGCTGCTGCTGAAAATTGCTGAAGACGCTGCAGAACTGACAGACGCT 656  
Qy 151 aagaacacgccttatgtatattatgcagttataacttgcacaaatttggcgttagaa 210  
Db 657 GGTAGCCAAAGAACCCGCTCAATTGCTTTGCAAGACTACTTGAACCAACCAACCTAAGAAC 716  
Qy 211 cccgaatttgtgttcagacgcgaagaatttaatgcagcggtgagtcctgtlaaaggt 270  
Db 717 CAGTTCACAAAGAGTCCCGAAGCTTAAGAGGTTGAATTAGACGCTGGAAGAAATTGAACCT 776  
Qy 271 gatgtgaagctctttgtgtgtgtatgtacatatgctataacttataatccccattt 330  
Db 777 GCTGAAAAGGAAGCTTTACGTTCCACCAACCAAGGTCAGAAAGCTCAAAATCTAAGAAATTGG 836  
Qy 331 tatgcagagggcaaatlttagcctgtcctaagataagatagttacacgcgttagtga 390  
Db 837 AAGACCAAGAGACTACTTGGAAATTTCATGCCACTTTGTTGAATCTACACTGGAAGAAC 896  
Qy 391 actacatactcaaaaaagcgacaaacccgctcagaagcggtgtgtgtgtgtg 446  
Db 897 TTCGGTGACAGAAACACACACAGCAAGAAACACTTCAACCAACCGTGTGTGTAG 952

## RESULT 12

AE001127/c 10976 bp DNA BCT 15-DEC-1997  
LOCUS AE001127/c  
DEFINITION Borrelia burgdorferi (section 13 of 70) of the complete genome.  
ACCESSION AE001127 AE000783  
VERSION AE001127.1 GI:2688047  
KEYWORDS  
SOURCE Lyme disease spirochete.  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 10976)

## REFERENCE

REFERENCE  
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,  
Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi  
JOURNAL Nature 390 (6660), 580-586 (1997)  
MEDLINE 98065943  
REFERENCE 2 (bases 1 to 10976)  
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,  
Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

FEATURES  
source  
location/Qualifiers  
1. .109/6  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
107. .823  
/gene="BB0158"  
107. .823  
/gene="BB0158"  
/note="similar to PID:1063419 GB:AB000790 percent  
identity: 37.28; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="antigen, S2, putative"  
/protein\_id="AAC66550.1"  
/db\_xref="GI:2688051"  
/translation="MVFPRYKHLLEIMLPMMLSCAPFKKPOSHODSNTGKSIDEX  
LHLISGKTSNKKLPIIINSNDYTWIKTKAMTILGSDGKEIPFKNKFGSIYISPKM  
DKKSYASLILFETTKNGDDEYIEIDKPTAGSTLEKNSLAVENSQEGRYTA  
IPEGLMSDEIKNAKRLTYKNGHWMYMLADLVKNKRLYQETKIKYRISLNSKLIIEFLK  
EVLKENSILKDIAGLFDI"  
847. .1521  
/gene="BB0159"  
847. .1521  
/gene="BB0159"  
/note="hypothetical protein; identified by glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0159"  
/protein\_id="AAC66555.1"  
/db\_xref="GI:2688056"  
/translation="MSIKKFIITLILSLAKNSPSENEINIPENENYIVKENIKTEIK  
KLKQFLASVDVAISOPIELADNGPIKELGISEYSFINVFSKIGSALISPDLS  
NEASKKYIKLEFLSPDKGNFTINOLSSITGSKQSKKELADAVSPGLPRESISKI  
IAEYKDNMYIILAIVENNINKETEKERIRINPKIYNDQKRLHFKSNQIKKE  
PIPIE"  
1569. .2687  
/gene="BB0160"  
1569. .2687  
/gene="BB0160"  
/note="similar to SP:P29743 PID:396388 GB:U00096  
PID:1790487 percent identity: 29.89; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="alanine racemase (alt)"  
/protein\_id="AAC66549.1"  
/db\_xref="GI:2688050"  
/translation="MYNNKTGMSNKTIIININLNLNLEHNLINKIGEKELVATLKGA  
YGHGLINFEKFLSKNINYPGLSINEDAKTLKIDKNIKILMYIKVKKKEIKNLKFE  
LVEGVSDEYELIEKCALOKNKIKVHLKIDIGNRRGKIDDLALETATYONSKPL  
ELPGLCSHLPSTENETOKOKEOFLPELTKOKNHKPKVHISNSGIIINXKNO  
ENMVRPGLILGYCQSLKNNKRPALNFKPYLSFSVITIKVKKGEKISYSITPAKE  
DMKILIPIGFDGIPONISNDFYPLINKKCKIKGKCOMLTIIVEIPDLKVKGSK  
VEIYSEKLSIDEMSEKFSRSHLELLCNIGKEYEKRYLD"  
/complement(2679. .4343)  
/gene="BB0161"  
/complement(2679. .4343)  
/gene="BB0161"  
/note="hypothetical protein; identified by glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0161"  
/protein\_id="AAC66554.1"  
/db\_xref="GI:2688055"  
/translation="MIDLLYNNKIYONRVLSNLFMLFIIFISAPLSFEKEYVYEFAD  
YSNYSIYKYEITKESEEEYGFYIDRNNNLILVDYIGKLSILMPSFNAQIKI  
ESRSKSEIRIFLNGLAKNNNGVYIEHIEWMDGRKININENYNSNEMVRDKYVSY  
YHLYDNRESYSVFNESGIDIKDLNNVYFKISYDHRKPKTYLIDENGYIYRSK  
NGTGFRLTYDNNHLIKEEYDKSAIVASNFSVATKIYHVDVGKYEILNYYINN  
NLTPENNVAIYREYFPRKDCYKREYNNNNMLTISONGYAMKKTIFYIYONNEKR

gene  
CDS  
11NSNKNKNYNDIPVYEEKYEIMDNFKGAIYSKYKDYDNFYLIENIPDKNFNL  
IADVGVMYIYRSYDKEGFLQAEHFGGFVNPIDIDFEGSVKYSYDNGNMIISKKN  
YSKDGLVADCNDFEYELIYEDKONRLISOKNFGSIGLODDIHGVSYAKYEYNNKKG  
ISKQSHGPDRLRDVGGSFYKKNIYDSRGLIDPQRKDSIGNLI"  
4357. .4524  
/gene="BB0162"  
4357. .4524  
/gene="BB0162"  
/note="hypothetical protein; identified by glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0162"  
/protein\_id="AAC66553.1"  
/db\_xref="GI:2688054"  
/translation="MAKISKNNQVSGKRLISRMGKIIMSKFENGKIKHYAECQTS  
KNTARKPRDL"  
/complement(4551. .6299)  
/gene="BB0163"  
/complement(4551. .6299)  
/gene="BB0163"  
/note="hypothetical protein; identified by glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0163"  
/protein\_id="AAC66552.1"  
/db\_xref="GI:2688053"  
/translation="MKKLEPIYMSKSEVDILISQGISSDPQKSLIKTLYSYKKEAI  
DTSFIEHSVKIRLKGCELDLFRKLVEKRGKGYTNRAVATGEPILPSIILTDASYS  
DEYVYSIDDLFLKTYIGIELFESSLTVKNIESSLVKNFETITADSDNFKFEKTS  
QKEILYWDKSLIIPSSRALYEFYPAKIFFNASNNVADVPRIRKGLKIDYEKLE  
EKGNIIGASKISLIGIEIKDEVLGKKGITFSKNYFLEFENLPIKNEKOLEEKAI  
EAVKIDESIKALEKTELERKKSIDNEEPFGILKYVSEGLNDPENLIVANNYPRFSKI  
HDILIKITQVKNFPLIKPRAPFELEEKYAKLANSIRASTIRKIIISFGDPSIIF  
TEEGLEVLSLVKSYRIMNEYCLKNNKFLDLIVSHYRITCTQSVQVHLKDLIVEA  
AAEFKNSNELPLRYKIRKLDYSLINAAVKDLSFKRYVLFTGKHQSYKEALSRIDF  
RAGLKMSGSGLEDSNERVKRQLEIKRQREEIEQVRNSFPKKAQOVASKSYTKO  
EODEAMVKEFSDRIOKK"  
/complement(6380. .7393)  
/gene="BB0164"  
/complement(6380. .7393)  
/gene="BB0164"  
/note="similar to GB:L77117 SP:057556 PID:1590872 percent  
identity: 34.47; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="Na+/Ca+ exchange protein, putative"  
/protein\_id="AAC66548.1"  
/db\_xref="GI:2688049"  
/translation="MFLPVEKYEYIOHLIOFVYVGRGIFILYIGNLLKSSGVIATYL  
KYPILILGYTYISFESAPLETSVAAKFKNEIVSNVIGSNIITIMLALPLAGFE  
LKIRKIDFKRLKSFMEFLMLLILLSLDFDSYFSPFVYVYRFSFSILFLSYLL  
LFYKEKEKVASLSENFQEGASNLNOSFNIPEFNLSFVISMVFLGSLIADVGALYI  
ANNVNVSEKILGIIILVAFGTSVPELVVSLFAIIRKESDIAFGNTIGSINIFNGFLA  
SSSFPRLIDQIYILDSIMWFITLVFLVAKFGVGRGSLIFLLIFYINLMLE  
NRY"  
/complement(7394. .9238)  
/gene="BB0165"  
/complement(7394. .9238)  
/gene="BB0165"  
/note="hypothetical protein; identified by glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0165"  
/protein\_id="AAC66551.1"  
/db\_xref="GI:2688052"  
/translation="MYMKFIYVWVYIFPFSVKKVYSIYSLNDEEFPKYSGLFVHKGF  
LSKNVNGKITKQVANGINSRWYPPYKILPSPSTISYEDVYSSSSFLTSSNLYVSYD  
YSKNRKLKLVGIDKFNAGVITSSAFSDQDKRIAGTAIHGILYLSVNGAISPKNNRL  
IPIYILGAGYDIISALIEFSKRETNLTYFSSGVYDIFLISQSGFIKKISIPPKQOI

Y	93	tgctcaacctatgltgtgccaataattgtcgaagtagacgccaaagcaatcaacggttaa	152
QY	33	agctcttcgattatgagcagatgagctcaacgctgcatacgctatgccaattcgtcgtga	92
Db	117	AGACTCACTGTCCTTTGTTTAATGGAATAAATTTTGTGAAGCAAAAGAAAAAATAGCTC	176
Y	93	tgctcaacctatgltgtgccaataattgtcgaagtagacgccaaagcaatcaacggttaa	152

QY	153	gaacacccctatglatatcagcaggtatcaacttgcacaaatttggcgtgaacc	212
DB	237	AAACATGCTTCTGGAACCCCTGTAAGGTTCAAGGCTTCAACAAGTAAGTAATTAAC	296
QY	213	cgaatttgctggtcagacgccaaagaattcaatca	249
DB	297	AGTTCTGCTGATTATTAACACAGTAACCTTGAAAGCA	333
RESULT	14		
LOCUS	AF077664	873 bp	DNA
DEFINITION	Borrelia burgdorferi arthropod-specific outer surface protein B (ospB) gene, partial cds.		BCF 03-NOV-1998
ACCESSION	AF077664		
VERSION	AF077664.1	GI:3386498	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 11:12:59 ; Search time 403.72 Seconds  
(without alignments)  
336.506 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543  
Sequence: 1 atgaatacttaaacact.....ggcctcatttgctttaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	7.6	1762	1 Q90739	B31 Osp-A/antigen
2	40.4	7.4	18613	1 V74423	Staphylococcus aur
3	40.2	7.4	110000	1 X20248_07	Continuation (8 of
4	39.4	7.3	846	1 T43302	OspB soluble varia
5	39.4	7.3	891	1 Q90713	B. burgdorferi str
6	39.4	7.3	891	1 T43315	OspB variant #1 co
7	39.4	7.3	1141	1 Q90737	B31 outer surface
8	39.4	7.3	1180	1 Q90735	B31 outer surface
9	39.4	7.3	1324	1 Q90738	B31 outer surface
10	39.4	7.3	1363	1 Q90736	B31 outer surface
11	39.4	7.3	1915	1 Q04651	OspA and OspB-enco
12	39.4	7.3	1916	1 V00681	Borrelia burgdorfe
13	39.4	7.3	5898	1 V00677	DNA plasmid VR211
14	39.4	7.3	53585	1 X20251	Borrelia burgdorfe
15	38.4	7.1	3323	1 T84223	DNA encoding a glu
16	38.4	7.1	3323	1 V53546	DNA encoding 2 sta
17	36.4	6.7	400	1 V78384	Staphylococcus aur
18	35.6	6.5	3794	1 V00680	Plasmid pTRH46 con
19	35.4	6.5	1214	1 V74424	Staphylococcus aur
20	35	6.4	19446	1 V52184	Streptococcus pneu
21	34.6	6.4	1349	1 X14172	H. pylori GHP0 639
22	33.6	6.3	4373	1 T72719	Pyruvate:flavodoxi
23	33.6	6.2	110000	1 X20248_08	Continuation (9 of
24	33.4	6.2	4023	1 X13023	Enterococcus faeca
25	33.2	6.1	1015	1 Q40695	B. burgdorferi str
26	32.4	6.0	2455	1 V26363	Moraxella catarrha
27	32.4	6.0	2896	1 Q24135	Vector cloning
28	32.2	5.9	110000	1 V21209_06	Continuation (7 of
29	31.8	5.9	730	1 V98768	DNA encoding a S.
30	31.8	5.9	2531	1 V42988	Streptococcus pneu
31	31.4	5.8	110000	1 V21209_00	Methanococcus jann
32	31.4	5.8	110000	1 V21209_01	Continuation (2 of
33	31.2	5.7	584	1 V62140	HSV-2 strain SB5 C
34	31	5.7	2696	1 Q41061	Ap Serotype 7 60KD

C	35	33	5.7	12438	1	X13110	Enterococcus faeca
C	36	31	5.7	235033	1	V57926	Hereditary haemoch
C	37	31	5.7	237326	1	V57903	Hereditary haemoch
C	38	30.8	5.7	2407	1	V26361	Moraxella catarrina
C	39	30.8	5.7	3762	1	T73217	ApXica gene, Prepa
C	40	30.8	5.7	4042	1	T04132	Helicobacter pylori
C	41	30.8	5.7	7528	1	X12992	Enterococcus faeca
C	42	30.8	5.7	110000	1	T58840_2	Continflation (3 of
C	43	30.6	5.6	2278	1	T08098	Enterohaemorrhagic
C	44	30.6	5.6	2472	1	V40652	Rat fatty acid am1
C	45	30.6	5.6	5857	1	V52161	Streptococcus pneu

## ALIGNMENTS

RESULT	ID	Score	Length	DB ID	Description
090739	1	7.6	1762	1 Q90739	B31 Osp-A/antigen
AC	30-JUL-1996	(first entry)			
DE	B31 Osp-A/antigen P41 (122-234)/Osp-C fusion gene.				
KW	Strain: B31; antigen; antigenic domain; protein; Osp-C				
KW	treatment; diagnosis; infection; vaccine; Lyme borreliosis;				
KW	immunodiagnostic assay; antibody; T-cell reactivity;				
KW	outer surface protein; Osp-A; antigen P41; fusion; ds.				
OS	Borrelia burgdorferi.				
FH	Key				
FT	mat_peptide	1.1762			
FT	mat_peptide	1.1762			
PN	NO9512676-A1.				
PD	11-MAY-1995.				
PF	27-OCT-1994; U12352.				
PR	01-NOV-1993; US-148191.				
PR	29-APR-1994; US-235836.				
PA	(ASU- ) ASSOC UNIVERSITIES INC.				
PI	Dunn JI, Luft BJ;				
DR	WPI; 95-215034/28.				
DR	P-PSDB: R75746.				
PT	Chimeric protein comprising 2 or more antigenic Borrelia				
PT	polypeptide(s) - useful in a vaccine against Lyme borreliosis and in				
PT	immuno:diagnostic assays				
PS	Claim 43; Fig 37; 200p; English.				
CC	The present sequence encodes a fusion protein comprising outer				
CC	surface protein A (Osp-A), antigen P41 (122-234) and Osp-C				
CC	sequences, from the B. burgdorferi strain B31. Using chemical or				
CC	enzymatic methods, peptide fragments of the parent proteins were				
CC	prep'd, and analysed by western blot to assess their ability to				
CC	bind different anti-Osp-A/P41/Osp-C monoclonal antibodies. The				
CC	information obtd. was used to locate antigenic domains in the				
CC	proteins, the epitopes of which were mapped with the aid of site				
CC	directed mutagenesis. Identical analyses were performed on a				
CC	selection of antigens purified from a variety of B. burgdorferi				
CC	strains, the results from which were utilised in the prep'n. of a				
CC	pool of antigenic Borrelia polypeptides, and corresponding				
CC	polynucleotides. Fusion proteins (i.e. the above protein)				
CC	comprising 2 or more antigenic Borrelia polypeptides, that do				
CC	not naturally occur in the same protein, can be used in the				
CC	treatment and diagnosis of Borrelia infections, i.e. as a				
CC	vaccine against Lyme borreliosis, in immunodiagnostic assays to				
CC	detect anti-Borrelia antibodies or to measure T-cell reactivity.				
SQ	Sequence 1762 BP; 700 A; 291 C; 343 G; 428 T;				
Query Match	7.6%	Score 41;	DB 1;	Length 1762;	
Best Local Similarity	49.3%	Pred. No. 0.0039;			
Matches 107;	Conservative 0;	Mismatches 110;	Indels 0;	Gaps 0;	
QY	33	agctctcgtatgtgagatgagtgatgacgctgcatgcatgcaattcgtcga	92		
DB	114	AGACGCACTGCTTTGTTTAAATGTAATAAATTTTGAAGCAAGAAAATAGCTC	173		
QY	93	tgctaacacctatgtgtgtgcacaaattgtgcaagtagaagcgaacgaacacggtaa	152		

```

DB 174 TGGCAAAATGATTAGACAAATTCATCGATTGAACCTTAAAGAACTTCGATGA 233
QY 153 gaacacccgtatggtattatgcaggtatataacttgacccaatttgcgctagaacc 212
DB 234 AAGCAATGGTTCGGAGAACCTTGAAAGCTTGACAAAGAGTAAGTAATAATTAA 293
QY 213 cgaattgttgctcagacgcaccaagaattaatga 249
DB 294 AGTTCTGCTGATTAAACACAGTAACTTAGAACGA 330

RESULT 2
V74423
ID V74423 standard; DNA; 18613 BP.
AC V74423;
DC 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SRQ ID #112.
KM Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc-feature 901..960
   /tag= a
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 2701..2760
   /tag= b
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 4501..4560
   /tag= c
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 6301..6360
   /tag= d
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 8101..8160
   /tag= e
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 9901..9960
   /tag= f
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 11701..11760
   /tag= g
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 13501..13560
   /tag= h
   /note= "these bases represent a line of missing text in
           the sequence listing in the specification. They
           are included to maintain the nucleotide numbering
           given in the specification for this DNA sequence"
FT misc-feature 15301..15360

```

```

FT /tag= i
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 17101..17160
FT /tag= j
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
PI WPI; 97-374922/35.
DR Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 641-651; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 18613 BP; 5990 A; 2753 C; 3477 G; 5782 T;

Query Match 7.4%; Score 40.4; DB 1; Length 18613;
Best Local Similarity 51.7%; Pared No. 0.017;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 91 gatgtcaaccctatgtgtgccaataatgtlcaagtagacgcacgaacaaatcaacggt 150
   ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5362 GATGCTCTACGATTTCATGTCGAAGACATGCGGTACACTTTATACCAACGTTGAATTG 5421
QY 151 aagaacacccgtatggtattatgcaggtatataacttgacccaatttgcgctaga 210
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5422 GAAAACAAAGGCGTGACTTTATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5481
QY 211 ccgaattgttgctcagacgcaccaagaattaatgaattatgcagcggtgtgtcttaaga 268
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5482 AAGCTTGATTGATGACGACTGTAAATGATGTAACCAAGCTTGTAACCGGAACAAG 5539

RESULT 3
X20248_07
Continuation (8 of 10) of X20248 from base 700001 (Borrelia burgdorferi polynucleotide)
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End
WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000

```



WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match	7.4%;	Score 40.2;	DB 1;	Length 110000;
Best Local Similarity	49.3%;	Pred. No. 0.043;		
Matches 105;	Conservative 0;	Mismatches 108;	Indels 0;	Gaps 0

OY 163 tatgtattatcagcgltataacttgaaccaaaatttgcgtagaacccgaatttgt 222  
| | | | | | | | | | | | | | |  
DB 46445 TTTCCTAAAAAATATTCTTTTAAATATGCTCTCATTAATGTCTGAAGAATGTTTTCGTT 46500

Qy 223 ggllcagacgcaccaagaatttaatgcaggcgtgagtcctgtataaagltgatgtgaagtc 282  
 Db 46505 AGAATAAAAGCATTAAGCTTATGTGATTATGAGAAAACCTTAAAGAAGCGGTAAACATATT 46564

Qy 283 ttctggtgcctatgycacatatcgcataactcatcaatacccatltaatgccaaagc 342  
||||||| | | | | | | | |  
Db 4655 attcgtcgaatttccaagattctctctgstatattgaanccaagattctaaggc 4662

QY 343 aaattagcattgctaagaactaaagcatgtt 375  
||| ||| ||| | ||| |||  
Db 46625 AAAAAAGGAATTACTTTTCAAAAAATTAATTTT 466

RESULT 4  
T43302  
TD T43302 standard. DNA. 846 bp

AC	143502;	
DT	10-FEB-1997	(first entry)
DE	OSPB soluble variant coding sequence	
KW	OSPB; OSPB; outer surface protein A; OSNB; OSPB	

KM  
KM  
KM  
KM  
KM  
ds.

Borrelia burgdorferi; Lyme disease; relapsing fever; dermatological disorder; arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein.

OS	BOLLEIDA DARGUOLLEI
PN	US5571718-A.
PD	05-NOV-1996.
PF	08-SEP-1992: 941523.

PA (ASUY-) ASSOC UNIVERSITIES INC.  
PI Barbour AG, Dunn JY;

DR	p-PSDB; W08096,
PT	Soluble recombinant forms of Borrelia lipo:proteins - useful for vaccine prodn. for treatment of Lyme disease
PT	

CC for the soluble recombinant variant of the *Borrelia burgdorferi* outer  
CC proteins of the invention. This sequence represents the coding sequence  
CC T43301-T43303 represent coding sequences for soluble recombinant *Borrelia*  
CC

burgdorferi spirochete. Borrelia spirochetes are responsible for a variety of human disorders including Lyme borreliosis, and relapsing fevers. The spirochete is transmitted to humans and animals through the

neurological and other pathogenic disorders in an infected host. This sequence is used to create recombinant host cells, and the encoded OsgD protein can be isolated from the cytosol of one of these cells without

CC antigens for the production of vaccines against Lyme disease. The  
CC recombinant proteins can also be used in immunoassays and other  
CC diagnostic screening methods to detect the presence of antibodies against  
CC antigens. The use of the recombinant proteins in the production of

Query Match	Score	DB 1: Length	DB 2: Length
846 BP;	367 A;	134 C;	150 G;
Sequence			195 T;

Query Match	7.3%;	Score 39.4;	DB 1;	Length 846;
Best Local Similarity	48.8%;	Pred. No. 0.0087;		
Matches 106;	Conservative 0;	Mismatches 111;	Indels 0;	Gaps 0

QY 33 agcttctcgattatgycgatgagtcgaagcgcgcacatcagctatgycaatctcgtcga 92  
 || || || || | || || | || || | || || |  
 Db 117 AGACCCAGTGTCTTCTTTTAAGCGTAATAAATTTTGTGAAGCAAGAAAAAATACCTC 176

QY 93 tgcTcaacctatgttggTgccaatatgTlcaagtagcgcacgaacaatcaacgtaa 152  
| | | | | | | | | | | | | | | |  
Db 177 CGGCAATATGATTTTAAGAGCACAACATTGATCAGTTGACTTAAAGAACCTCCGATTA 236

Oy 153 gaacaccgcctatgtgtaattatgcaggtataactttgaccaaatlttgcgtagaac 212  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 237 AACCATGTTCTGGAAccCTTGAAGTTCAAAGcCTGACAGAgaGTAAATTAAC 296

```

QY      213 cgaatttgttgcagacgcgaagaatttaatgca 249
          | | | | | | | | | | | | | |
Db      297 AGTTCTGCTGATTAAACACAGTAACCTTGAAGCA 333

```

RESULT	5
Q90713	

AC Q90713;  
DT 31-JUL-1996 (first entry)  
DE B. burgdorferi strain B31 outer surface protein B (OspB-B31) DNA  
ID 1011 outer surface protein B (OspB-B31) DNA

KM chlamyeric protein; treatment; diagnosis; infection; vaccine  
 KW Lyme borreliosis; immunodiagnostic assay; antibody;  
 T-cell reactivity; chimeric; ds.  
 OS *Borrelia burgdorferi*

PN	W09512676-A1	
FT	cds	Location/Qualifiers
FT	Key	1..891
FT		/*tag= a

PD 11-MAY-1995. U12352.  
PF 27-OCT-1994; U12352.  
PR 01-NOV-1993; US-148191.  
PR 29-APR-1994; US-235836.

PI Dunn JJ, Luft BJ  
DR WPI; 95-215034/28  
DR P-PSDB; R75726.

PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
 PT immuno:diagnostic assays  
 PS Example 1; Fig 11; 200pp; English.

CC surface protein B (OspB-B31). Using chemical or enzymatic methods  
CC peptide fragments of OspB-B31 were prepd., and analysed by westerns  
CC blot to assess their ability to bind different anti-OspB monoclonals

CC domains in OspB-B3j, the epitopes of which were mapped with the  
CC aid of site directed mutagenesis. Identical analyses were performed  
CC on a selection of Osp purified from a variety of *B. burgdorferi*

CC pool of antigenic *Borrelia* polypeptides, and corresponding  
CC polynucleotides. Chimeric proteins comprising 2 or more antigenic  
CC *Borrelia* polypeptides, that do not naturally occur in the same

CC infections, i.e. as a vaccine against Lyme borreliosis, in  
CC immunodiagnostic assays to detect anti-Borrelia antibodies or to  
CC measure T-cell reactivity.

Query Match	7.38;	Score 39.4;	DB 1;	Length 891;
Post Total Cited/Ref	48	00	0000	

33	agccttcctgattgacgatgatagcgaacgcgcacatcagcatggaattctgcga	92
Matches	106; Conservative	0; Mismatches 111; Indels 0; Gaps 0;

Db 162 AGACTCAGTGTCTTTGTTAATGCTAATAAATTTTGTGAAGCAAGAAAAAATAGCTC 222

93    tgc<sup>+</sup>taac<sup>+</sup>cctatgtt<sup>+</sup>gttgc<sup>+</sup>caaatgt<sup>+</sup>lcaagtag<sup>+</sup>acgccaag<sup>+</sup>caatca<sup>+</sup>cgttaa<sup>+</sup> 154

```
Db 222 CGGCAATATGATTTAAGACCAACATTCAGTTGACTTAAGAACTCCGATTA 281
OY 153 gaacacgcctatgattgattatgcaggtataaacttgacaaatttggcgtagaacc 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 AAACAATGGTTCGGAACCCCTTAAGGCTTCACAAAGAGTAAGTAATAATTAAAC 341
OY 213 cgaattgtgtgtgcagcgccaaagaatttaatga 249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 342 AGTTTCTGCTGATTTAACACAGTAACCTTAAGAAACA 378
```

## RESULT 6

```
T43315
ID T43315 standard; DNA; 891 BP.
AC T43315;
DT 10-FEB-1997 (first entry)
DE OspA: outer surface protein A; Borrelia; variable major protein 7;
KW OspA; OspB: outer surface protein B; Borrelia; variable major protein 7;
KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;
KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;
KW ds.
OS Synthetic.
PN US5571718-A.
PD 05-NOV-1996.
PF 08-SEP-1992; 941523.
PR 21-DEC-1990; US-632072.
PR 08-SEP-1992; US-941523.
PA (ASUV-) ASSOC UNIVERSITIES INC.
PI Barbour AG, Dunn JT;
DR WPI: 96-505409/50.
PT Soluble recombinant forms of Borrelia lipo:proteins - useful for
PT vaccine prodn. for treatment of Lyme disease
PS Example 8; Column 39-42; 49pp; English.
CC This sequence represents the coding sequence for a soluble recombinant
CC Borrelia outer surface protein B (OspB) variant proteins of the
CC invention. OspB is a surface lipoprotein of the B. burgdorferi
CC spirochete. Borrelia spirochetes are responsible for a variety of human
CC disorders including Lyme borreliosis, and relapsing fevers. The
CC spirochete is transmitted to humans and animals through the bite of a
CC tick, and can cause serious dermatological, arthritic, neurological and
CC other pathogenic disorders in an infected host. This sequence is used to
CC create recombinant host cells, and the encoded OspB protein can be
CC isolated from the cytosol of one of these cells without the use of
CC detergent. The encoded recombinant proteins can be used as antigens for
CC the production of vaccines against Lyme disease. The recombinant
CC proteins can also be used in immunoassays and other diagnostic screening
CC methods to detect the presence of antibodies against Borrelia
CC lipoproteins in the sera of infected patients.
SQ Sequence 891 BP; 380 A; 137 C; 160 G; 214 T;
```

Query Match 7.3%; Score 39.4; DB 1; Length 891;  
Best Local Similarity 48.8%; Pred. No. 0.0089;  
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
OY 33 agcttcgttattgtagatgagtgctacagctgcacatcagctatgcaattctgtga 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 AGACTCAGTGTCTTTTAATGTAATAAATTTTGTGAAGCAAAAGAAAATATGCTC 221
OY 93 tgcacaacctatgtgtgtgccaataatgtgcaagtagacgcgaacaatacaagtgtaa 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 CGGCAATATGATTTAAGACCAACATTCAGTTGAACTTAAGAACTCCGATTA 281
OY 153 gaacacgcctatgattgattatgcaggtataaacttgacaaatttggcgtagaacc 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 AAACAATGGTTCGGAACCCCTTAAGGCTTCACAAAGAGTAAGTAATAATTAAAC 341
OY 213 cgaattgtgtgtgcagcgccaaagaatttaatga 249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 342 AGTTTCTGCTGATTTAACACAGTAACCTTAAGAAACA 378
```

```
RESULT 7
OY0737
ID OY0737 standard; DNA; 1141 BP.
AC OY0737;
DT 30-JUL-1996 (first entry)
DE B31 outer surface protein (Osp-A)/antigen P41 (140-234) fusion gene.
KW Strain; B31; antigen; antigenic domain; protein;
KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;
KW immunodiagnostic assay; antibody; T-cell reactivity;
KW outer surface protein; Osp-A; antigen P41; fusion; ds.
OS Borrelia burgdorferi.
FH Key Location/Qualifiers
FT mat_peptide 1..1141
FT /tag= a
```

```
PN WO9512676-A1.
PD 11-MAY-1995.
PF 27-OCT-1994; U12352.
PR 01-NOV-1993; US-148191.
PR 29-APR-1994; US-235836.
PA (ASUV-) ASSOC UNIVERSITIES INC.
PI Dunn JT, Luft BJ;
DR WPI: 95-215034/28.
DR P-PSDB; R75744.
PT Chimeric protein comprising 2 or more antigenic Borrelia
PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in
PT immuno:diagnostic assays
PS Claim 43; Fig 35; 200pp; English.
CC The present sequence encodes a fusion protein comprising outer
CC surface protein A (Osp-A) and antigen P41 (140-234) sequences, from
CC the B. burgdorferi strain B31. Using chemical or enzymatic methods,
CC peptide fragments of the parent proteins were prep'd, and analysed
CC by western blot to assess their ability to bind different
CC anti-Osp-A/P41 monoclonal antibodies. The information obtd. was used
CC to locate antigenic domains in the proteins. The epitopes of which
CC were mapped with the aid of site directed mutagenesis. Identical
CC analyses were performed on a selection of antigens purified from a
CC variety of B. burgdorferi strains, the results from which were
CC utilised in the prepn. of a pool of antigenic Borrelia polypeptides,
CC and corresponding polynucleotides. Fusion proteins (i.e. the
CC above protein) comprising 2 or more antigenic Borrelia polypeptides,
CC that do not naturally occur in the same protein, can be used in the
CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine
CC against Lyme borreliosis, in immunodiagnostic assays to detect
CC anti-Borrelia antibodies or to measure T-cell reactivity.
SQ Sequence 1141 BP; 449 A; 206 C; 221 G; 265 T;
```

Query Match 7.3%; Score 39.4; DB 1; Length 1141;  
Best Local Similarity 48.8%; Pred. No. 0.0099;  
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
OY 33 agcttcgttattgtagatgagtgctacagctgcacatcagctatgcaattctgtga 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 AGACTCAGTGTCTTTTAATGTAATAAATTTTGTGAAGCAAAAGAAAATATGCTC 173
OY 93 tgcacaacctatgtgtgtgccaataatgtgcaagtagacgcgaacaatacaagtgtaa 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 CGGCAATATGATTTAAGACCAACATTCAGTTGAACTTAAGAACTCCGATTA 233
OY 153 gaacacgcctatgattgattatgcaggtataaacttgacaaatttggcgtagaacc 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 AAACAATGGTTCGGAACCCCTTAAGGCTTCACAAAGAGTAAGTAATAATTAAAC 293
OY 213 cgaattgtgtgtgcagcgccaaagaatttaatga 249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 AGTTTCTGCTGATTTAACACAGTAACCTTAAGAAACA 330
```

## RESULT 8

```
OY0735
ID OY0735 standard; DNA; 1180 BP.
AC OY0735;
DT 30-JUL-1996 (first entry)
```

DE B31 outer surface protein (Osp-A)/antigen P41 (122-234) fusion gene.  
 KW Strain; B31; antigen; antigenic domain; protein;  
 KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;  
 KW immunodiagnostic assay; antibody; T-cell reactivity;  
 KW outer surface protein; Osp-A; antigen P41; fusion; ds.  
 OS Borrelia burgdorferi.  
 FH Key location/Qualifiers  
 FT mat\_peptide 1..1180  
 FT /\*tag= a  
 PN MO9512676-A1.  
 PD 11-MAY-1995.  
 PE 27-OCT-1994; US-12352.  
 PR 01-NOV-1993; US-148191.  
 PR 29-APR-1994; US-235836.  
 PA (ASUY-) ASSOC UNIVERSITIES INC.  
 PI Dunn JF, Luft BJ;  
 DR WPI: 95-215034/28.  
 DR P-PSDB: R75742.  
 PT Chimeric protein comprising 2 or more antigenic Borrelia  
 PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
 PT immuno:diagnostic assays  
 PS Claim 43; Fig 33; 200pp; English.  
 CC The present sequence encodes a fusion protein comprising outer  
 CC surface protein A (Osp-A) and antigen P41 (122-234) sequences, from  
 CC the B. burgdorferi strain B31. Using chemical or enzymatic methods,  
 CC peptide fragments of the parent proteins were prep'd., and analysed  
 CC by western blot to assess their ability to bind different  
 CC anti-Osp-A/P41 monoclonal antibodies. The information obt'd. was used  
 CC to locate antigenic domains in the proteins, the epitopes of which  
 CC were mapped with the aid of site directed mutagenesis. Identical  
 CC analyses were performed on a selection of antigens purified from a  
 CC variety of B. burgdorferi strains, the results from which were  
 CC utilised in the prep'n. of a pool of antigenic Borrelia polypeptides,  
 CC and corresponding polynucleotides. Fusion proteins (i.e. the  
 CC above protein) comprising 2 or more antigenic Borrelia polypeptides,  
 CC that do not naturally occur in the same protein, can be used in the  
 CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine  
 CC against Lyme borreliosis, in immunodiagnostic assays to detect  
 CC anti-Borrelia antibodies or to measure T-cell reactivity.  
 SQ Sequence 1180 BP; 466 A; 215 C; 224 G; 275 T;

Query Match 7.3%; Score 39.4; DB 1; Length 1180;  
 Best Local Similarity 48.8%; Pred. No. 0.01;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agctcttcgtataggcgatgagtgctaacgctgcacatcagctatgcgaattcgtcga 92  
 DB 114 AGACTCAGTGCTGTTGTTTAATGTAATTAATTTTGTAAAGCAAAATAATAGCTC 173  
 OY 93 tgcataccctatgtgtgtccaaattgtaagtaagcgcacaaatcaacgtaaa 152  
 DB 174 CGGCAAAATATGATTTTAAGACAAATGATGATGAGTAAGCAACTTCCGATTA 233  
 OY 153 gaacacgcctatggtattatgcaggtataacttgaccacaaatttggcgtagaac 212  
 DB 234 AAACAATGTTCTGGAAACCTTGAAGGTTCAAAAGCTGACAAAGTAAGTAATAATTAAC 293  
 OY 213 cgaattgtgttcagagcccaagaatttaatgca 249  
 DB 294 AGTTTCTGCTGATTTTAACACAGTAACCTTAGAAGCA 330

RESULT 9  
 ID Q90738 standard; DNA; 1324 BP.  
 AC Q90738;  
 DT 30-JUL-1996 (first entry)  
 DE B31 outer surface protein (Osp-A)/antigen P41 (140-295) fusion gene.  
 KW Strain; B31; antigen; antigenic domain; protein;  
 KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;  
 KW immunodiagnostic assay; antibody; T-cell reactivity;  
 KW outer surface protein; Osp-A; antigen P41; fusion; ds.  
 OS Borrelia burgdorferi.  
 FH Key location/Qualifiers  
 FT mat\_peptide 1..1363  
 FT /\*tag= a  
 PN MO9512676-A1.

OS Borrelia burgdorferi.  
 FH Key location/Qualifiers  
 FT mat\_peptide 1..1324  
 FT /\*tag= a  
 PN MO9512676-A1.  
 PD 11-MAY-1995.  
 PE 27-OCT-1994; US-12352.  
 PR 01-NOV-1993; US-148191.  
 PR 29-APR-1994; US-235836.  
 PA (ASUY-) ASSOC UNIVERSITIES INC.  
 PI Dunn JF, Luft BJ;  
 DR WPI: 95-215034/28.  
 DR P-PSDB: R75745.  
 PT Chimeric protein comprising 2 or more antigenic Borrelia  
 PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
 PT immuno:diagnostic assays  
 PS Claim 43; Fig 36; 200pp; English.  
 CC The present sequence encodes a fusion protein comprising outer  
 CC surface protein A (Osp-A) and antigen P41 (140-295) sequences, from  
 CC the B. burgdorferi strain B31. Using chemical or enzymatic methods,  
 CC peptide fragments of the parent proteins were prep'd., and analysed  
 CC by western blot to assess their ability to bind different  
 CC anti-Osp-A/P41 monoclonal antibodies. The information obt'd. was used  
 CC to locate antigenic domains in the proteins, the epitopes of which  
 CC were mapped with the aid of site directed mutagenesis. Identical  
 CC analyses were performed on a selection of antigens purified from a  
 CC variety of B. burgdorferi strains, the results from which were  
 CC utilised in the prep'n. of a pool of antigenic Borrelia polypeptides,  
 CC and corresponding polynucleotides. Fusion proteins (i.e. the  
 CC above protein) comprising 2 or more antigenic Borrelia polypeptides,  
 CC that do not naturally occur in the same protein, can be used in the  
 CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine  
 CC against Lyme borreliosis, in immunodiagnostic assays to detect  
 CC anti-Borrelia antibodies or to measure T-cell reactivity.  
 SQ Sequence 1324 BP; 521 A; 232 C; 249 G; 322 T;

Query Match 7.3%; Score 39.4; DB 1; Length 1324;  
 Best Local Similarity 48.8%; Pred. No. 0.011;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agctcttcgtataggcgatgagtgctaacgctgcacatcagctatgcgaattcgtcga 92  
 DB 114 AGACTCAGTGCTGTTGTTTAATGTAATTAATTTTGTAAAGCAAAATAATAGCTC 173  
 OY 93 tgcataccctatgtgtgtccaaattgtaagtaagcgcacaaatcaacgtaaa 152  
 DB 174 CGGCAAAATATGATTTTAAGACAAATGATGATGAGTAAGCAACTTCCGATTA 233  
 OY 153 gaacacgcctatggtattatgcaggtataacttgaccacaaatttggcgtagaac 212  
 DB 234 AAACAATGTTCTGGAAACCTTGAAGGTTCAAAAGCTGACAAAGTAAGTAATAATTAAC 293  
 OY 213 cgaattgtgttcagagcccaagaatttaatgca 249  
 DB 294 AGTTTCTGCTGATTTTAACACAGTAACCTTAGAAGCA 330

RESULT 10  
 ID Q90736 standard; DNA; 1363 BP.  
 AC Q90736;  
 DT 30-JUL-1996 (first entry)  
 DE B31 outer surface protein (Osp-A)/antigen P41 (122-295) fusion gene.  
 KW Strain; B31; antigen; antigenic domain; protein;  
 KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;  
 KW immunodiagnostic assay; antibody; T-cell reactivity;  
 KW outer surface protein; Osp-A; antigen P41; fusion; ds.  
 OS Borrelia burgdorferi.  
 FH Key location/Qualifiers  
 FT mat\_peptide 1..1363  
 FT /\*tag= a  
 PN MO9512676-A1.

PD 11-MAY-1995. 012352.  
 PF 27-OCT-1994; 012352.  
 PR 01-NOV-1993; 05-148191.  
 PR 29-APR-1994; US-235836.  
 PA (ASUY-) ASSOC UNIVERSITIES INC.  
 PI Dunn JF, Luft BJ.  
 PI WPI: 95-215034/28.  
 DR P-PSDB; R75743.  
 PT Chimeric protein comprising 2 or more antigenic Borrelia  
 PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
 PT immuno:diagnostic assays  
 PS Claim 43: Fig 34: 200pp: English.  
 CC The present sequence encodes a fusion protein comprising outer  
 CC surface protein A (OspA) and antigen P41 (122-293) sequences, from  
 CC the B. burgdorferi strain B31. Using chemical or enzymatic methods,  
 CC peptide fragments of the parent proteins were prep., and analysed  
 CC by western blot to assess their ability to bind different  
 CC anti-OspA/P41 monoclonal antibodies. The information obd. was used  
 CC to locate antigenic domains in the proteins, the epitopes of which  
 CC were mapped with the aid of site directed mutagenesis. Identical  
 CC analyses were performed on a selection of antigens purified from a  
 CC variety of B. burgdorferi strains, the results from which were  
 CC utilised in the prep. of a pool of antigenic Borrelia polypeptides,  
 CC and corresponding polynucleotides. Fusion proteins (i.e. the  
 CC above protein) comprising 2 or more antigenic Borrelia polypeptides,  
 CC that do not naturally occur in the same protein, can be used in the  
 CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine  
 CC against Lyme borreliosis, in immunodiagnostic assays to detect  
 CC anti-Borrelia antibodies or to measure T-cell reactivity.  
 SQ Sequence 1363 BP; 538 A; 241 C; 252 G; 332 T;

Query Match 7.3%; Score 39.4; DB 1; Length 1363;  
 Best Local Similarity 48.8%; Pred. No. 0.011;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agctcttcgtatattgagcagatgagctcaagctccatcaactatgcaattctgctga 92  
 DB 114 AGACTCAGTGTCTTGTGTTAATGTAATTAATTTTGTAAACCAAGAAAATAATGCTC 173  
 OY 93 tgcacacctatgtgtggtccaaattgctcaagtacagtcgaagcgaacatacaacgtaa 152  
 DB 174 CGGCAAAATATGATTTAAGGCAACAATTTGATCAGTTGAACCTTAAAGAACTTCCGATTA 233  
 OY 153 gaacacgctatagtattatgcaggtatataactttgacccaatttggcgtagaacc 212  
 DB 234 AACAATATGTTCTGGAACCTTGAAGGTTCAAGCCTGACAAAGTAAGTAATAATTAAAC 293  
 OY 213 cgaatttgtgttcagacgccaagaatttaatga 249  
 DB 294 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 330

RESULT 11  
 ID 004651 standard; DNA: 1915 BP.  
 AC 004651;  
 DE 02-OCT-1990 (first entry)  
 DE OspA and Ospb-encoding sequence of Borrelia burgdorferi  
 DE Borrelia burgdorferi; Lyme disease; active immunisation;  
 KM passive immunisation; OspA; Ospb; ss.  
 OS Borrelia burgdorferi.  
 FH Key Location/Qualifiers  
 FT rbs 1..138  
 FT /tag= a  
 FT /label= OspA RBS  
 FT cds 151..972  
 FT /tag= b  
 FT /label= OspA structural gene  
 FT -10\_signal 102..107  
 FT /tag= c  
 FT /label= -10 region of P1  
 FT -10\_signal 108..113

FT /tag= d  
 FT /label= -10 region of P2  
 FT 80..85  
 FT /tag= e  
 FT /label= -35 region of P1  
 FT 87..92  
 FT /tag= f  
 FT /label= -35 region of P2  
 FT 972..977  
 FT /tag= g  
 FT /label= Ospb RBS  
 FT cds 982..1873  
 FT /tag= h  
 FT /label= Ospb structural gene

WO9004411-A.  
 PD 03-MAY-1990.  
 PE 24-OCT-1989; W00248.  
 PR 24-OCT-1988; DK-005902;  
 PA (SYMB-) Symbicom AB.  
 PI Bergstrom S, Barbour AG, Magnarelli LA;  
 PI WPI: 90-163873/21.  
 DR P-PSDB; R05028, R05029  
 PT New immunologically active fractions of Borrelia burgdorferi -  
 PT and derived antibodies and DNA coding sequences for active and  
 PT passive immunisation against and diagnosis of Lyme disease.  
 PS Disclosure; 112pp: English.  
 CC OspA and Ospb genes are part of the same operon located on a linear  
 CC plasmid of B. burgdorferi and encode outer membrane proteins. They  
 CC are separated by 12 bases within which is located the RBS of Ospb.  
 CC There are two putative promoters (P1 and P2) upstream of OspA. Ospb  
 CC include two closely spaced direct repeats of a 12-mer sequence  
 CC (start-13 and 29) and a 14-mer palindromic sequence (start-123) which  
 CC surrounds the P1 and P2 -10 regions.  
 SQ Sequence 1915 BP; 811 A; 279 C; 337 G; 488 T;

Query Match 7.3%; Score 39.4; DB 1; Length 1915;  
 Best Local Similarity 48.8%; Pred. No. 0.012;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agctcttcgtatattgagcagatgagctcaagctccatcaactatgcaattctgctga 92  
 DB 1143 AGACTCAGTGTCTTGTGTTAATGTAATTAATTTTGTAAACCAAGAAAATAATGCTC 1202  
 OY 93 tgcacacctatgtgtggtccaaattgctcaagtacagtcgaagcgaacatacaacgtaa 152  
 DB 1203 CGGCAAAATATGATTTAAGGCAACAATTTGATCAGTTGAACCTTAAAGAACTTCCGATTA 1262  
 OY 153 gaacacgctatagtattatgcaggtatataactttgacccaatttggcgtagaacc 212  
 DB 1263 AACAATATGTTCTGGAACCTTGAAGGTTCAAGCCTGACAAAGTAAGTAATAATTAAAC 1322  
 OY 213 cgaatttgtgttcagacgccaagaatttaatga 249  
 DB 1323 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 1359

RESULT 12  
 ID V00681 standard; DNA: 1916 BP.  
 AC V00681;  
 DE 08-JUN-1998 (first entry)  
 DE Borrelia burgdorferi ospA and ospb gene region.  
 DE Borrelia; antigen; outer surface protein; OspA; Ospb; Lyme disease;  
 KM DNA vaccine; immunisation; VR2210; VR2211; vector; ss.  
 OS Borrelia burgdorferi strain B31.  
 FH Key Location/Qualifiers  
 FT RBS 139..144  
 FT /tag= a  
 FT /label= CDS  
 FT 152..973  
 FT /tag= b  
 FT /label= sig\_peptide  
 FT 152..199  
 FT /tag= c

```

FT mat_peptide 200..970
FT /*tag= d
FT 973..979
FT RBS /*tag= e
FT 983..1873
FT CDS /*tag= f
FT 983..1027
FT sig_peptide /*tag= g
FT 1028..1870
FT mat_peptide /*tag= h
FT
FT W09747197-A1.
FT 18-DEC-1997.
FT 03-JUN-1997; U09439.
FT 14-JUN-1996; US-663998.
FT (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
FT (UYTE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
FT (VICA-) VICAL INC.
FT Barbour AG, Garner KR, Huebner RC, Liang X, Luke CJ,
FT Norman JA;
FT WPI: 98-051938/05.
FT
FT DR P-PSDB; W26771 AND W37131.
FT
FT PT Plasmid for expressing Borrelia antigen in eukaryotic cells - used
PT as vaccines to protect against Lyme disease and for production of
PT antigens, themselves used in vaccines
PT
PS Example 1: Fig 7A-H; 80pp; English.
CC This nucleotide sequence comprises the ospA and ospB gene region
CC of Borrelia burgdorferi. The ospA and ospB genes respectively
CC code for outer surface proteins (Osp) OspA (see W37131) and OspB
CC (see W26771) of B. burgdorferi. The ospA and ospB genes have
CC been used in the construction of VR2210 (see V00676) and VR2211
CC (see V00677). DNA plasmids suitable for transfection of e.g. baby
CC hamster kidney and human melanoma cells. Plasmid DNA encoding at
CC least one Borrelia genospecies antigen is disclosed and claimed.
CC The genospecies may be B. burgdorferi, Borrelia garinii and/or
CC Borrelia afzelii. The antigen can be OspA and/or OspB and/or
CC The plasmid is used in vaccines to elicit a protective immune
CC response (both humoral and cellular) in hosts, human or animal,
CC susceptible to Lyme disease (claimed). It can also be used for the
CC production of a Borrelia antigen in vitro in eukaryotic cells
CC (claimed), and these antigens used as vaccines or immunogenic
CC compositions, or to produce monoclonal antibodies.
CC
SQ Sequence 1916 BP; 810 A; 280 C; 337 G; 489 T;

```

Query Match 7.3%; Score 39.4; DB 1; Length 1916;  
 Best Local Similarity 48.8%; Pred. No. 0.012;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

QY 33 agcttcctcgatattgagcagatgagtgctaacgctgcacatgcatatgcattctgcctga 92
DB 1144 AGACTCAGTGTCTTGTGTTAATGTAATAAATTTTGTAGCAAAAGAAAATAGCTC 1203
QY 93 tgcatacccatctgtgtgtgccaataatggtcgaagtagacgccaagcaatcaacggttaa 152
DB 1204 CGGCAAAATATGATTTAAGAGCAACATTCATCAGTTCACTTAAGAGAACTCCGATTA 1263
QY 153 gaacacccgtatggtattatgcaggtataacttgcacaaaatttggcgtagaacc 212
DB 1264 AAACAATGTTCTGGAAACCTTGAAGGTTCAAAAGCCTGACAAAGATAAGTAAATTAAC 1323
QY 213 cgaatttggttcagagcgcacaaagaatttaagca 249
DB 1324 AGTTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 1360

```

```

OS Chimeric - Borrelia burgdorferi strain B31.
OS Chimeric - Homo sapiens.
FT Key location/Qualifiers
FT primer_bind Complement (2..19)
FT /*tag= a
FT /note= "binding site of forward primer used to
FT amplify and insert OspB coding sequence"
FT primer_bind 835..852
FT /*tag= b
FT /note= "binding site of reverse primer used to
FT amplify and insert OspB coding sequence"
FT
FT W09747197-A1.
FT 18-DEC-1997.
FT 03-JUN-1997; U09439.
FT 14-JUN-1996; US-663998.
FT (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
FT (UYTE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
FT (VICA-) VICAL INC.
FT Barbour AG, Garner KR, Huebner RC, Liang X, Luke CJ,
FT Norman JA;
FT WPI: 98-051938/05.
FT
FT DR Plasmid for expressing Borrelia antigen in eukaryotic cells - used
FT as vaccines to protect against Lyme disease and for production of
FT antigens, themselves used in vaccines
FT
PS Example 1: Fig 4A-D; 80pp; English.
CC This is the nucleotide sequence of VR2211, a DNA plasmid suitable
CC for transfection of e.g. baby hamster kidney and human melanoma
CC cells. VR2211 contains DNA encoding the Borrelia burgdorferi
CC outer surface protein B (OspB). It was made by ligating: an OspB
CC fragment (see V00680) amplified from pTRH46 (see V00685-86); a
CC tissue plasminogen activator 5'-untranslated region/leader peptide
CC sequence amplified from nckvminb (see V00678); and pscI- and
CC XbaI-digested plasmid VR1012. Plasmid DNA encoding at least one
CC Borrelia genospecies antigen is disclosed and claimed. The
CC genospecies may be B. burgdorferi, Borrelia garinii and/or
CC Borrelia afzelii. The antigen can be OspA and/or OspB and/or
CC The plasmid is used in vaccines to elicit a protective immune
CC response (both humoral and cellular) in hosts, human or animal,
CC susceptible to Lyme disease (claimed). It can also be used for the
CC production of a Borrelia antigen in vitro in eukaryotic cells
CC (claimed), and these antigens used as vaccines or immunogenic
CC compositions, or to produce monoclonal antibodies.
CC
SQ Sequence 5898 BP; 1643 A; 1402 C; 1380 G; 1473 T;

```

Query Match 7.3%; Score 39.4; DB 1; Length 5898;  
 Best Local Similarity 48.8%; Pred. No. 0.021;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

QY 33 agcttcctcgatattgagcagatgagtgctaacgctgcacatgcatatgcattctgcctga 92
DB 118 AGACTCAGTGTCTTGTGTTAATGTAATAAATTTTGTAGCAAAAGAAAATAGCTC 177
QY 93 tgcatacccatctgtgtgtgccaataatggtcgaagtagacgccaagcaatcaacggttaa 152
DB 178 CGGCAAAATATGATTTAAGAGCAACATTCATCAGTTGAACCTTAAGGAACCTCCGATTA 237
QY 153 gaacacccgtatggtattatgcaggtataacttgcacaaaatttggcgtagaacc 212
DB 238 AAACAATGTTCTGGAAACCTTGAAGGTTCAAAAGCCTGACAAAGATAAGTAAATTAAC 297
QY 213 cgaatttggttcagagcgcacaaagaatttaagca 249
DB 298 AGTTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 334

```

RESULT 13  
 V00677  
 ID V00677 standard; DNA: 5898 BP.  
 AC V00677;  
 DT 08-JUN-1998 (first entry)  
 DE DNA plasmid VR2211 containing Borrelia burgdorferi OspB gene.  
 KW Borrelia; antigen; outer surface protein A; OspB; Lyme disease;  
 KW DNA vaccine; immunisation; VR2211; vector; ss.

RESULT 14  
 X20251/c  
 ID X20251 standard; DNA: 53585 BP.  
 AC X20251;  
 DT 04-MAY-1999 (first entry)  
 DE Borrelia burgdorferi polynucleotide sequence #4.  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;





**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 11:05:58 ; Search time 278.23 Seconds  
(without alignments)  
247,802 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543  
Sequence: 1 atgaacaatttaaacact.....gcgcctcattgcttttaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 segs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCBUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.4	7.3	846	1	US-07-941-523-21 Sequence 21, Appl
2	39.4	7.3	891	1	US-07-941-523-19 Sequence 19, Appl
3	39.4	7.3	1959	2	US-08-137-175A-1 Sequence 1, Appl
4	39.4	7.3	5952	3	US-08-663-998-2 Sequence 2, Appl
5	31	5.7	2696	1	US-07-961-522-1 Sequence 1, Appl
6	31	5.7	2696	1	US-08-217-438-1 Sequence 1, Appl
7	31	5.7	2696	2	US-08-321-978-1 Sequence 1, Appl
8	31	5.7	2696	2	US-08-710-584-1 Sequence 1, Appl
9	31	5.7	246240	3	US-08-724-394A-20 Sequence 20, Appl
10	31	5.7	246240	3	US-08-724-394A-21 Sequence 21, Appl
11	31	5.7	246240	3	US-08-724-394A-22 Sequence 22, Appl
12	30.8	5.7	2520	4	US-08-450-351-1 Sequence 1, Appl
13	30.8	5.7	3762	5	US-08-772-370A-3 Sequence 3, Appl
14	30.8	5.7	4042	2	US-08-200-232-1 Sequence 1, Appl
15	30.8	5.7	4042	2	US-08-200-232-1 Sequence 1, Appl
16	30.8	5.7	4042	2	US-08-200-232-1 Sequence 1, Appl
17	30.8	5.7	4042	2	US-08-200-232-1 Sequence 1, Appl
18	30.8	5.7	4042	2	US-08-200-232-1 Sequence 1, Appl
19	30.6	5.6	2278	2	US-08-258-188-1 Sequence 1, Appl
20	30.6	5.6	2278	2	US-08-258-188-1 Sequence 1, Appl
21	30.6	5.6	2278	2	US-08-258-188-1 Sequence 1, Appl
22	29.8	5.5	1903	1	US-07-961-522-3 Sequence 3, Appl
23	29.8	5.5	1903	1	US-08-217-438-3 Sequence 3, Appl
24	29.8	5.5	1903	2	US-08-321-978-3 Sequence 3, Appl
25	29.8	5.5	1903	3	US-08-710-584-3 Sequence 3, Appl
26	29.4	5.4	660	3	US-08-625-377-1 Sequence 1, Appl
27	29.2	5.4	1529	4	US-08-477-451-17 Sequence 17, Appl

C	28	29.2	5.4	1529	4	US-08-477-451-21	Sequence 21, Appl
	29	29.2	5.4	19932	5	US-08-477-451-25	Sequence 25, Appl
	30	28.8	5.3	2949	4	US-08-433-522A-1	Sequence 1, Appl
	31	28.8	5.3	3294	1	US-08-409-995-1	Sequence 1, Appl
	32	28.6	5.3	5319	2	US-08-169-927-1	Sequence 1, Appl
	33	28.4	5.2	2802	1	US-08-215-805A-79	Sequence 79, Appl
	34	28.4	5.2	3848	1	US-08-215-805A-1	Sequence 1, Appl
	35	28.2	5.2	2329	4	US-08-982-232-13	Sequence 13, Appl
	36	28	5.2	211	2	US-08-435-684A-54	Sequence 54, Appl
	37	28	5.2	211	4	US-08-934-877A-54	Sequence 54, Appl
	38	28	5.2	825	4	US-08-441-857-1	Sequence 1, Appl
	39	28	5.2	1020	4	US-08-441-857-5	Sequence 5, Appl
	40	28	5.2	1302	2	US-08-425-299A-1	Sequence 1, Appl
	41	28	5.2	1452	2	US-08-187-785-2	Sequence 2, Appl
	42	27.8	5.1	1506	1	US-08-149-105-2	Sequence 2, Appl
	43	27.8	5.1	1506	1	US-08-317-847-2	Sequence 2, Appl
	44	27.8	5.1	1988	2	US-08-469-202-26	Sequence 26, Appl
	45	27.8	5.1	1988	4	US-08-484-434C-33	Sequence 33, Appl

## ALIGNMENTS

RESULT 1  
US-07-941-523-21  
Sequence 21, Application US/07941523  
Patent No. 5571718  
GENERAL INFORMATION:  
APPLICANT: Dunn, John J  
TITLE OF INVENTION: Cloning and Expression of Borrelia  
TITLE OF INVENTION: Lipoproteins  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 01730  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/941,523  
FILING DATE: 19920908  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BNL90-01A  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 846 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-941-523-21

Query Match 7.3%, Score 39.4, DB 1, Length 846;  
Best Local Similarity 48.8%, Pred. No. 0.0029;  
Matches 106, Conservative 0, Mismatches 111, Indels 0, Gaps 0;  
Cy 33 agctcttcgtatgtggcgaatgctcgaacgcctcgaacgtatgcaattcgcgcga 92  
Db 117 AGACTCAAGTGTCTTGTGTTATGTAATTAATTTTGTATGCAAGAAATAATGCTC 176

Qy	93	tgccaaacccatgctgctgctgcacaattgtgcagatgagacgcgaagaacaaatcaagttaa	152
Db	177	CGGCAAAATTAAGATTTTAAAGACCAACAATGATGACGGTTGAACCTTAAAGGAACCTTCCGATTA	236
Qy	153	gaacaccgcctatcgtratacttatcgagttataactttgacacaatttgcgtagaaacc	212
Db	237	AAACAAATGGTTCCTGGAAACCTTGAAGGTTTAAAGCCTGACAAAGATTAAGTAATAATTTAAC	296
Qy	213	cgaaattcgtctgcagacgcgaagaagatttaatgca	249
Db	297	AGTTTCGCTGATTTAAACACAGTAACCTTAGAAGCA	333

```

RESULT 2
US-07-941-523-19
: Sequence 19, Application US/07941523
Patent No. 5571718
GENERAL INFORMATION:
APPLICANT: Dunn, John J
APPLICANT: Barbour, Alan G
TITLE OF INVENTION: Cloning and Expression of Borrelia
TITLE OF INVENTION: Lipoproteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
City: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 01730
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/941,523
FILING DATE: 19920908
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BNL90-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 891 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-941-523-19

```

Query Match	7.3%	Score	39.4	DB	1	Length	891
Best Local Similarity	48.8%	Pred. No.	0.0029				
Matches	106	Conservative	0	Mismatches	111	Indels	0
				Gaps			0

[illegible]

Qy 213 cgaatttgttggltcagaagccaaagaatttaatgca 249  
| | | | | | | | | | | | | | | | | |  
Db 342 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 378

```

RESULT      3
US-08-137-175A-1
: Sequence 1, Application US/08137175A
: Patent No. 577095
:
: GENERAL INFORMATION:
: APPLICANT: BARBOUR, Alan G.
: APPLICANT: BERGSTROEM, Sven
: APPLICANT: HANSSON, Lennart
: TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND
: TITLE OF INVENTION: PROPHYLAXIS
: NUMBER OF SEQUENCES: 22
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEWMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/137,175A
: FILING DATE: 26-OCT-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCY/US92/08972
: FILING DATE: 22-OCT-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: COOPER, Iyer P.
: REGISTRATION NUMBER: 28,005
: REFERENCE/DOCKET NUMBER: BARBOUR-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SRO ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1959 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Borrelia burgdorferi
: STRAIN: B31 (ATCC 35210)
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 123..142
: OTHER INFORMATION: /function= "Primer"
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 584..607
: OTHER INFORMATION: /function= "Primer"
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 776..794
: OTHER INFORMATION: /function= "Primer"
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 806..817
: OTHER INFORMATION: /function= "Primer"
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 119..940
: OTHER INFORMATION: /product= "ospA"

```



[illegible]

```

RESULT 6
US-08-217-438-1
: Sequence 1, Application US/08217438
: Patent No. 5521072
:
: GENERAL INFORMATION:
: APPLICANT: Potter, Andrew A.
: APPLICANT: Gerlach, Gerald F.
: APPLICANT: Willson, Philip J.
: APPLICANT: Rossi-Campos, Analia
:
: TITLE OF INVENTION: ACTINOBACILLUS PLEUROPEIMONIAE
: TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS AND USES THEREOF
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Reed & Robins
: STREET: 635 Bryant Street
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94301
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/217,438
: FILING DATE: 22-MAR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Robins, Roberta L.
: REGISTRATION NUMBER: 33,208
: REFERENCE/DOCKET NUMBER: 9001-0015.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-617-8999
: TELEFAX: 415-327-3231
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2696 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 333..1973
:
: US-08-217-438-1

```

Query Match	5.7%;	Score 31;	DB 1;	Length 2696;
Best Local Similarity	47.0%;	Pred. No. 2.1;		
Matches 131;	Conservative	0;	Mismatches 145;	Indels 3;
				Gaps 1;

Oy	235	aaagaatttaacgaagcgagagtcctgttaaaagygatgaagtcctttggtcctat	294
Db	1662	ACAGAAATTTCATGTCATTTTGGCACAACAAAATTTTAAGTGGTAAGTATTTCATTA	1721
Oy	295	ggcacatcgcatalaactcatalcaatbacccatttalcgcaagggcaaatltagcat	354
Db	1722	GGTAAATCTCTGTGTTTACCGTCAGATGCAAAAATTTGAATGATGAATGTTTACTGGCAAA	1781
Oy	355	gctaagactaaagtatgtgtaccagcgtaatgtaacatactcaacaagaaggaac	414
Db	1782	GCTAAACCTCACATTAAGAGCTTCGCTCTAGATTCACGTAAGTTCACGTTATGACAAATG	1841
Oy	415	aaaaccagctagacgagcggtgttggtgttggtcttaaacattagcaaatgtggcgt	474
Db	1842	AAATTTAACATGTAGACAGTACTGCTGTGCTCTTAATGTCCAACGCGA---GCACAGCTT	1898
Oy	475	gaagcaagctacacactctatcaagaagttgcacaagca	513
Db	1899	GGCGACAAATTCACCACTTAATATCAGAAATTTGGCAGTGTA	1937

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	5.7%;	Score 31;	DB 2;	Length 2696;
Best Local Similarity	47.0%;	Pred. No. 2.1;		
Matches 131;	Conservative	0;	Mismatches 145;	Indels 3;
				Gaps 1;

QY 235 aaagaatttaagcagcggtgagctcgttaaagaagtgtgaagcttttggtcttat 294  
 Db 1662 ACAGAAATTGATTTGATTTGGCAACAAAATTTAAGTGTAGTTATTGATTAATAAAC 1721  
 QY 295 ggcacatacgtcataactcaatcatcaatcaccatcttatgccaaggccaattagcatt 354  
 Db 1722 GGTGTAAATCCTGCTGTATACCGTAGATGCAAAAATTCATGCTATGCTTTACTGCGCAA 1781  
 QY 355 gctaaagactaagtagatgtgtaccagcgtaagtcaactacatactcaacaaagcgac 414  
 Db 1782 GCTAAACCTCAGATGAGGCTTCGCTGTAGATTGAGTTCACGTTATGAGAAATGTG 1841  
 QY 415 aaaccagcctagcagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474  
 Db 1842 AAATTTAAGCATGTAGAGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1898  
 QY 475 gaagcaagctacaactatctatcagaagaatgccaatgca 513  
 Db 1899 GCGGACAAATTCACCATTAATCAGAAATGCGAGTGT 1937

## RESULT 8

; Sequence 1, Application US/08710584  
 ; Patent No. 5876725  
 ; GENERAL INFORMATION:  
 ; APPLICANT: POTTER, ANDREW A.  
 ; APPLICANT: GERLACH, GERALD F.  
 ; APPLICANT: WILSON, PHILIP J.  
 ; APPLICANT: ROSSI-CAMPOS, AMALIA  
 ; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: ROBERTA L. ROBINS  
 ; STREET: 285 HAMILTON AVE, SUITE 200  
 ; CITY: PALO ALTO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 94301  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/710,584  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/321,978  
 ; FILING DATE: 12-OCT-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: ROBINS, ROBERTA L.  
 ; REGISTRATION NUMBER: 33,208  
 ; REFERENCE/DOCKET NUMBER: 9001-0015.02  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 327-3400  
 ; TELEFAX: (415) 327-3231  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2696 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS 333..1973  
 ; LOCATION: 333..1973  
 ; US-08-710-584-1

Query Match 5.7%; Score 31; DB 3; Length 2696;  
 Best Local Similarity 47.0%; Pred. No. 2.1; Mismatches 145; Indels 3; Gaps 1;  
 Matches 131; Conservative 0;

QY 235 aaagaatttaagcagcggtgagctcgttaaagaagtgtgaagcttttggtcttat 294  
 Db 1662 ACAGAAATTGATTTGATTTGGCAACAAAATTTAAGTGTAGTTATTGATTAATAAAC 1721  
 QY 295 ggcacatacgtcataactcaatcatcaatcaccatcttatgccaaggccaattagcatt 354  
 Db 1722 GGTGTAAATCCTGCTGTATACCGTAGATGCAAAAATTCATGCTATGCTTTACTGCGCAA 1781  
 QY 355 gctaaagactaagtagatgtgtaccagcgtaagtcaactacatactcaacaaagcgac 414  
 Db 1782 GCTAAACCTCAGATGAGGCTTCGCTGTAGATTGAGTTCACGTTATGAGAAATGTG 1841  
 QY 415 aaaccagcctagcagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474  
 Db 1842 AAATTTAAGCATGTAGAGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1898  
 QY 475 gaagcaagctacaactatctatcagaagaatgccaatgca 513  
 Db 1899 GCGGACAAATTCACCATTAATCAGAAATGCGAGTGT 1937

## RESULT 9

; Sequence 20, Application US/08724394A  
 ; Patent No. 5872237  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Feder, John N.  
 ; APPLICANT: Krommal, Gregory S.  
 ; APPLICANT: Laufer, Peter M.  
 ; APPLICANT: Ruddy, David A.  
 ; APPLICANT: Thomas, Winston  
 ; APPLICANT: Tsuchihashi, Zenta  
 ; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
 ; STREET: Two Embarcadero Center, 8th Floor  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94111-3834  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/724,394A  
 ; FILING DATE: 01-OCT-1996  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Filts, Renee A.  
 ; REGISTRATION NUMBER: 35,136  
 ; REFERENCE/DOCKET NUMBER: 017957-000100  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-576-0200  
 ; TELEFAX: 415-576-0300  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 246240 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-20

Query Match 5.7%; Score 31; DB 3; Length 246240;  
Best Local Similarity 51.0%; Pred. No. 15;  
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

16 acactatgacgagtcacgtctcttcgtatgagcgaagtcgaagtcacgcacgc 75  
DB 75473 ACACACTTGCACTACAACTTTGAGATTATTCGTCGCTGACCTACCTCCACATTATC 75532

76 tatggcaatcgcgcgagtcgaacctatgtgtgccaatggtcgaagtagacgcc 135  
DB 75533 AGAGCCACATTTCTTTATGGAATATTATGTTGTGCAAAAGTATTGCGGTCTTTGCC 75592

136 aagcaatcaacgctaagaacac 158  
DB 75593 ATTAAAGTAAAGCAAAAACACAC 75615

RESULT 10  
US-08-724-394A-21

Sequence 21, Application US/08724394A  
Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Krommal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESS: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-21

Query Match 5.7%; Score 31; DB 3; Length 246240;  
Best Local Similarity 51.0%; Pred. No. 15;  
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

16 acactatgacgagtcacgtctcttcgtatgagcgaagtcgaagtcacgcacgc 75  
DB 75473 ACACACTTGCACTACAACTTTGAGATTATTCGTCGCTGACCTACCTCCACATTATC 75532

76 tatggcaatcgcgcgagtcgaacctatgtgtgccaatggtcgaagtagacgcc 135  
DB 75533 AGAGCCACATTTCTTTATGGAATATTATGTTGTGCAAAAGTATTGCGGTCTTTGCC 75592

136 aagcaatcaacgctaagaacac 158  
DB 75593 ATTAAAGTAAAGCAAAAACACAC 75615

RESULT 11  
US-08-724-394A-22

Sequence 22, Application US/08724394A  
Patent No. 5872237

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Krommal, Gregory S.

APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston

APPLICANT: Tsuchihashi, Zenta

APPLICANT: Wolff, Roger K.

TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESS: TOWNSEND and TOWNSEND and CREW LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,394A

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.

REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 017957-000100

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200

TELEFAX: 415-576-0300

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 246240 base pairs

TYPE: nucleic acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..246240

OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-22

Query Match 5.7%; Score 31; DB 3; Length 246240;  
Best Local Similarity 51.0%; Pred. No. 15;  
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

[illegible]

```

RESULT# 12
US-08-450-351-1
: Sequence 1, Application US/08450351
: Patent No. 5981213
:
: GENERAL INFORMATION:
: APPLICANT: Hansen, Eric J.
: APPLICANT: Helminen, Merja E.
: APPLICANT: Maciver, Isobel
: TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO
: TITLE OF INVENTION: USEFUL ANTIGENS OF MORAXELLA CATARRHALIS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/450,351
: FILING DATE: 25-MAY-1995
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Highlander, Steven L.
: REGISTRATION NUMBER: 37,642
: REFERENCE/DOCKET NUMBER: AMCC:019
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000
: TELEFAX: (512) 474-7577
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2520 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
US-08-450-351-1

```

	Query Match	5.7%	Score 30.8	DB 4	Length 2520
	Best Local Similarity	42.8%	Pred. No. 2.3		
	Matches 155	Conservative 0	Mismatches 207	Indels 0	Gaps 0
QY	91 gatgtcaaacctatggtgtgcacaaattgtgtcaagtagaagccaaagcaatacaagt	150			
DB	1408 GAAGCTGTAACCCCTAGAGGTTTAGACCTGGTTTTCAAAACCAAGAAAAACCATGCT	1467			
QY	151 aagaacaccgcgtatagtattatgtcaggtataacttgcacaaattttgcgtagaa	210			
DB	1468 GGCATTATCTAGAAAGCGGTTAACCAAAATCANTGACTTTACATCAATATACAGCGTGGCT	1527			
QY	211 cccgaatttgtgtgtcagaagccaaagaatttaatgcagcgtgtgtcgttaaaagt	270			
DB	1528 GTTGAACCGTTTGTGACATTTAAAGCTATGACGGCTTAAAGATGGTGAAGAAACCATCACTCAAC	1587			

OY	271	gatgtgaagctcttttggigtactatgbacatatcgctaataccttaacaatcaaccacatt	330
Db	1588	CCAMGCTTTGGGGTGATTTTATGATGTCAATCCTAATCTTAGCGTCGCAGTAACCTAATC	1647
OY	331	tatgccaaaggccaatttaggcatttgctcaagaactcaaatgatgtatgccagcgatataca	390
Db	1648	TATGCCACTCCACGCCACCGCTTTGCTGATGCTATCTTAACCGGTGGCTTCCGTATAGC	1707
OY	391	actaatactcaaaaacaaagcgaaaaaacagactagtcagcgagtgttgttgcttt	450
Db	1708	GTTTGAGATTATTTGCTGATTAACGAAAGCAAAGAAAAGCACGCAATACGAGATTGTTTT	1767
OY	451	aa 452	
Db	1768	AA 1769	

```

1      RESULT 13
2      US-08-450-351-3
3      ; Sequence 3, Application US/08450351
4      ; Patent NO. 5981213
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Hansen, Eric J.
7      ; APPLICANT: Helminen, Merja E.
8      ; APPLICANT: Maciver, Isobel
9      ; TITLE OF INVENTION: METHODS AND COMPOSITIONS REL
10     ; TITLE OF INVENTION: USEFUL ANTIGENS OF MORAXELLA
11     ; NUMBER OF SEQUENCES: 4
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: Arnold, White & Durkee
14     ; STREET: P.O. Box 4433
15     ; CITY: Houston
16     ; STATE: Texas
17     ; COUNTRY: USA
18     ; ZIP: 77210
19
20     COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Floppy disk
22     ; COMPUTER: IBM PC compatible
23     ; OPERATING SYSTEM: PC-DOS/MS-DOS
24     ; SOFTWARE: Patentin Release #1.0, Version #1.308
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/08/450,351
27     ; FILING DATE: 25-MAY-1995
28
29     CLASSIFICATION: 435
30     ; ATTORNEY/AGENT INFORMATION:
31     ; NAME: Highlander, Steven L.
32     ; REGISTRATION NUMBER: 37,642
33     ; REFERENCE/DOCKET NUMBER: AMCY:019
34     ; TELECOMMUNICATION INFORMATION:
35     ; TELEPHONE: (512) 418-3000
36     ; TELEFAX: (512) 474-7577
37
38     INFORMATION FOR SEQ ID NO: 3:
39     ; SEQUENCE CHARACTERISTICS:
40     ; LENGTH: 2520 base pairs
41     ; TYPE: nucleic acid
42     ; STRANDEDNESS: single
43     ; TOPOLOGY: linear
44
45     FEATURE:
46     ; NAME/KEY: CDS
47     ; LOCATION: 100..2376
48
49     US-08-450-351-3

```

```
Query Match      5.7% ; Score 30.8 ; DB 4 ; Length 2520;
Best Local Similarity 42.8% ; Pred.No. 2.3 ;
Matches 155; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
```

Oy    91   gatgtcaaccctatgttggtgcacaanaattgltcaagtagcagccaaagaatacaagt   150  
       |||||    |||||    |||||    |||||    |||||    |||||

Db    1408   GAAGCGTGTACCGCCTGAGAAGTATTAGACTGGTTTTCAAAACCAAGAAAAAACCGATGCT   1467

Oy    151   aagaacaccgttatgttattatgcagtttaactttgaacaaaatttggcgtlagaa   210  
       |||||    |||||    |||||    |||||    |||||    |||||





Db 2891 AATGAAATCACCAGCAATGTGATACGGCCACTACCACTTA 2932

Search completed: May 20, 2000, 11:09:40  
Job time: 13402 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 10:35:43 ; Search time 5377.06 Seconds  
(without alignments)  
409.313 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543  
Sequence: 1 atgaactttaaaacact.....gcgcctcattgcttttaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

```
EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est13:*
60: gb_est134:*
61: gb_est135:*
62: gb_est136:*
63: gb_est137:*
64: gb_est138:*
65: em_est127:*
66: em_est128:*
67: em_est129:*
68: em_est130:*
69: gb_est139:*
70: gb_est140:*
71: gb_est141:*
72: gb_est142:*
73: gb_est143:*
74: gb_est144:*
75: em_est131:*
76: em_est132:*
77: em_est133:*
78: em_est134:*
79: gb_est145:*
80: gb_est146:*
81: gb_est147:*
82: gb_gss11:*
83: gb_gss12:*
84: gb_gss13:*
85: gb_gss14:*
86: em_gss11:*
87: em_gss12:*
88: em_gss13:*
89: em_gss14:*
90: gb_gss15:*
91: gb_gss16:*
92: gb_gss17:*
93: gb_gss18:*
94: gb_gss19:*
95: em_gss15:*
96: em_gss16:*
97: em_gss17:*
98: em_gss18:*
99: em_gss19:*
100: em_gss110:*
101: em_gss111:*
102: gb_gss110:*
103: gb_gss111:*
104: em_gss112:*
105: gb_gss112:*
106: gb_gss113:*
107: gb_gss114:*
108: gb_gss115:*
109: gb_gss116:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	39.4	7.3	1101	83	CNS0182P	AL108811 Drosophila
2	38.6	7.1	344	80	AW334656	AW334656 S38H1 AGS
3	38.6	7.1	1101	82	CNS00EAX	AL068861 Drosophila
4	36.6	6.7	587	103	AQ256569	AQ256569 nbdx0016B
5	36.4	6.7	479	90	AO826019	AO826019 HS_5442_A
6	36.4	6.7	1101	82	CNS008F6	AL052505 Drosophila
7	35.8	6.6	460	42	AL138233	AL138233 qd76e03.x
8	35.8	6.6	514	90	AO784640	AO784640 HS_3193_B
9	35.2	6.5	807	82	CNS00A7D	AL054816 Drosophila
10	35	6.5	1000	82	CNS00CQ	AL059446 Drosophila
11	34.8	6.4	407	74	AM201163	AM201163 se96f03.y
12	34.8	6.4	556	103	AQ176788	AQ176788 HS_3213_A
13	34.8	6.4	1086	82	CNS00YXK	AL069662 Drosophila
14	34.6	6.4	558	20	Z46831	Z46831 ATTTA393 Ra
15	34.6	6.4	593	60	AT194544	AT194544 fca5f09.y
16	34.6	6.4	801	107	AQ449605	AQ449605 500002D09
17	34.4	6.3	402	102	AQ124099	AQ124099 HS_3122_A
18	34.4	6.3	442	74	AM202121	AM202121 s12b08.y
19	34.4	6.3	460	62	AL1900885	AL1900885 sb95f12.y
20	34.4	6.3	464	81	AW397695	AW397695 sg83b10.y
21	34.2	6.3	530	103	AQ214218	AQ214218 HS_3011_A
22	34.2	6.3	797	82	CNS003F8	AL064634 Drosophila
23	34	6.3	347	21	R02601	R02601 ye/6a01.r1
24	33.8	6.2	374	32	AA375686	AA375686 EST88188
25	33.8	6.2	392	40	AA908681	AA908681 o104e04.s
26	33.8	6.2	559	35	AA557106	AA557106 948 l0blo
27	33.8	6.2	569	106	AQ384040	AQ384040 Rptc11-15
28	33.6	6.2	380	80	AV403461	AV403461 AV403461
29	33.6	6.2	395	47	AL153238	AL153238 SD03803.5
30	33.4	6.2	568	85	AO698184	AO698184 HS_5547_B
31	33.4	6.2	1039	83	CNS017AD	AL107791 Drosophila
32	33.2	6.1	365	47	AL1497040	AL1497040 fb59d08.y
33	33.2	6.1	493	85	AO718562	AO718562 HS_5509_B
34	33.2	6.1	521	61	AI827597	AI827597 w10g10.x
35	33.2	6.1	649	94	AQ003789	AQ003789 Cpg0394B
36	33	6.1	300	35	C37766	C37766 C37766 YqJ1
37	33	6.1	508	35	AA553216	AA553216 VK86B09.s
38	33	6.1	510	33	AA408349	AA408349 EST02828
39	33	6.1	532	81	AM398432	AM398432 EST298279
40	33	6.1	1101	82	CNS012RM	AL101332 Drosophila
41	32.8	6.0	381	38	AA746003	AA746003 nx87b01.s
42	32.8	6.0	485	107	AQ476428	AQ476428 CITBI-E1-
43	32.8	6.0	599	63	AI974558	AI974558 701554019
44	32.8	6.0	689	61	AI869704	AI869704 w198g02.x
45	32.6	6.0	431	25	N97920	N97920 1539c3 czap

## ALIGNMENTS

RESULT 1						
LOCUS	CNS0182P	1101 bp	DNA	GSS	26-JUL-1999	
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	AL108811					
VERSION	AL108811.1					
KEYWORDS	GI:5629115					
SOURCE	GSS:					
ORGANISM	fruit fly.					
REFERENCE	Drosophila melanogaster					
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
TITLE	1 (bases 1 to 1101)					
JOURNAL	Genoscope.					
COMMENT	Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a					

collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billard at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

## FEATURES

## source

1..1101  
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db\_xref="taxon:7227"  
/clone\_11b="DrosBAC"  
/clone="BACN37D10"  
/note="end : SP6"  
/note="end : 268 c 128 g 73 t 358 others

## BASE COUNT

274 a 268 c 128 g 73 t 358 others

## ORIGIN

## Query Match

7.3%; Score 39.4; DB 83; Length 1101;

Best local similarity 18.6%; Pred. No. 0.36;

Matches 66; Conservative 138; Mismatches 151; Indels 0; Gaps 0;

QY	189	tgaccacaaatttgccgtgaacccgaattgtgtcagacgccaagaatttaagc	248
DB	662	DGCGAAADAWFRGRRRAAARAAAADAKRAAAAAAADAAGGRRK	721
QY	249	aggcgagatcctgttaagaagtgatgtgaagcttgcgttaagcaatagcca	308
DB	722	WKGDKDGGKATKTAAMKGRKMDGTATAMWTMDTAMADTAKRAAARRR	781
QY	309	taactcacaatacccaattatgccaaaggaatagcatgctgaagtaagct	368
DB	782	DARKTRDGRARRRARRAAGGARARARARARARADRDMAAAAAA	841
QY	369	agatgtaccagccgcatgaatcacaatacgaagaagcgaagcctagc	428
DB	842	AAWTTTRDRHMDWDMDTAMWDATAAMWDARAARRRRRRARRAARDT	901
QY	429	aggcggtgtgtgtgtgtgttaaacattagcaaatgtgagcgttgaaagca	488
DB	902	DTKDRADATTDKTTTDDDDKAKRDMWMAKADGAKMWRDRAADAA	961
QY	489	ctatcacaagaagatgcaatgattgttgagcgtcatttgccttta	543
DB	962	DKMKDWGGRKGRGKDKKRMKDKKDDDKMTWTRDMWMTTRKMDMW	1016

RESULT 2						
LOCUS	AW334656/c	344 bp	mRNA	EST	31-JAN-2000	
DEFINITION	S38H1 AGS-1 Pneumocystis carinii f. sp. carinii cDNA 3', mRNA sequence.					
ACCESSION	AW334656					
VERSION	AW334656.1					
KEYWORDS	GI:6831117					
SOURCE	EST.					
ORGANISM	Pneumocystis carinii f. sp. carinii.					
REFERENCE	Pneumocystis carinii f. sp. carinii.					
AUTHORS	Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae; Pneumocystis.					
TITLE	1 (bases 1 to 344)					
JOURNAL	Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman,J.C., Kovacs,J. and Cushion,M.					
COMMENT	Expressed sequence tags from Pneumocystis carinii Unpublished (2000) On Jun 22, 1998 this sequence version replaced gi:3246719. Contact: Staben C School of Biological Sciences University of Kentucky 101 Morgan Building, University of Kentucky, Lexington, KY 40506-0225, USA Tel: 606 257 2161 Fax: 606 257 1717					

Email: [staben@pop.uky.edu](mailto:staben@pop.uky.edu).

FEATURES	Location/Qualifiers
source	1. .344

```

/organism="Pneumocystis carinii f. sp. carinii"
/db_xref="taxon:38081"
/clone_lib="AGS-1"
/lab_host="E. coli"
/note="Vector: Lambda ZAP II; Site.1: EcoRI; Site.2: XhoI
P. carinii organisms (3x106) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Total extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/projec/Pneumocystis/

```

BASE COUNT

Query Match	7.18;	Score 38.6;	DB 80;	Length 344;
Best Local Similarity	47.38;	Pred. No. 0.44;		
Matches 116; Conservative	0;	Mismatches 129;	Indels 0;	Gaps 0;

Qy	127	gtgaagcgcaagcaaatccaacggtgaagaaacacgcgttatgttatattatgaggttataac	186
Db	253	GAAAACGCTTATATTTTATAAAGAAATCTAAATACCTTATATAAACAAGCCGGTATGTATA	194
Qy	187	lttgaccaaaatttbtgcgtagaacccgaatttgttggttcagaagccaaagaattaat	246
Db	193	ATTGAACCAATCTTTGGTTTGTTCGCGAACCAACCTCATATTGCAAAAAATTTTAAT	134
Qy	247	gcagcgctgtgctcctgtaaaagtgatgtgaatctctttgttgcttatgtcacatatcgc	306
Db	133	GTGCTAAATCAATGTGCGAAAAGAAAGTAAAAAATCCATTTTATCTCAAGCATTTTAGAC	74
Qy	307	tataatctcaataaccccatctttagcgaagygcaaatlaagcatctgtcgaagctaaac	366
Db	73	GAAAATTCCTTCAAAAACAGAGTTTATATATCAAGTGAAGGAGAAATTTTCAAAAAAAA	14
Qy	367	gtaga	371
Db	13	AAAAA	9

RESULT	3
CNS00EAX/c	
LOCUS	
DEFINITION	
	CNS00EAX 1101 bp DNA GSS 04-JUN-1999
	Drosophila melanogaster genome survey sequence TET3 end of BAC #
	BACR28D24 of RPEC-98 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)

P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	source	location/qualifiers
	1..1101	
	/db_xref="Drosophila	melanogaster"
	/db_xref="taxon:7227"	
	/clone_lib="RRC1-98"	
	/clone="BACR28D24"	
	/note="end : 1E13"	
BASE COUNT	348 a	233 t
ORIGIN	175 c	200 others

Query Match	7.1%;	Score 38.6;	DB 82;	Length 1101;
Best Local Similarity	29.6%;	Pred. No. 0.62;		
Matches	85;	Conservative	66;	Mismatches 136;
			Indels	0;
			Gaps	0;

[illegible]

RESULT	4
A0256569	
LOCUS	A0256569 587 bp DNA
DEFINITION	nxbx001B19r CUGI Rice BAC Library Oryza sativa genomic clone nxbx001B19r, genomic survey sequence.
ACCESSION	A0256569
VERSION	A0256569.1 GI:3781051
KEYWORDS	GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa

REFERENCE	1 (bases 1 to 587)
AUTHORS	Wing, R. A. and Dean, R. A.
TITLE	A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL	Unpublished (1998)
COMMENT	Contact: Wing RA

```

/organism="Oryza sativa
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"

```

```
/cultivar="Nipponbare"  
/db_xref="taxon:4530"
```

```
/db_xref="taxon:4530"
```

```

/clone="hxb0016B19r"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBlotScript; Site 1: HindIII; Site 2:
HindIII. Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 kb providing 10.9 haploid genome equivalents.
The deep coverage allows the isolation a particular
sequence with a probability of 99.9%. Two high density
filters, each containing 18,432 clones (doubly spotted),
represent the whole library for colony screening."

BASE COUNT      183 a      128 c      112 g      164 t
ORIGIN

Query Match
Best Local Similarity 56.1%; Score 36.6; DB 103; Length 587;
Matches 69; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 309 taactcatcaatacccatcttaatgccaaggaataaggcattgctaagactaaagt 368
      ||||| ||| ||||| ||| ||||| ||||| ||||| ||||| |||||
DB 460 TCAATTAATTTGTTATCGCAATTAATGCCAGACATTAATGATGATTTCTAGACACAAAC 519
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 agatgtaccagccgaatgcaactatcaccacaagaagcgaacaaacagcctaac 428
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 520 AANAGCAATCAATCGCAATGACCCACATGACACGGAAGAACCCCTGATCAGCTAGC 579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 429 agg 431
      |||
DB 580 ACG 582

RESULT 5
A0826019 479 bp DNA GSS 27-AUG-1999
LOCUS HS-5442_A1-G03-SP6E-RPCT-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=1018 Col=5 Row=M, genomic survey sequence.
ACCESSION A0826019
VERSION A0826019.1 GI:5792081
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 479)
Mahtiras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahtiras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are delivered from the human BAC library RPCT-11. For BAC
library availability, please contact Pieter de Jong

```

```

(piet@ede.jong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1018 row: M column: 5
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 479.
Location/Qualifiers
1..479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1018 Col=5 Row=M"
/clone_lib="RPCT-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACe3.6; Genomic sequence of BAC ends"

BASE COUNT      151 a      83 c      98 g      145 t      2 others
ORIGIN

Query Match
Best Local Similarity 53.5%; Score 36.4; DB 90; Length 479;
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 82 aattctgtcatgtcacaacctatgttgcccaaatgtgtcaagtagacgccaaagca 141
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 335 ATTTCGCTTGAGAGCTCTCTTTATTTTGATCAAAATGGAAGAAAAAATTCACAGCAC 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 142 atcaagagtagaacaccgcttatgtattatgcacgtgtataacttgacaaatttt 201
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 395 ATGAAACAAAGAAAGGGGCAAAAGCTCTGTACTGTAAACAGACATCTTAGTAT 454
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 202 ggcgtagaaccgcgaattgtgt 223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 455 GGATCTAATTAAGGCTTATGC 476

RESULT 6
CNS008F6/6 1101 bp DNA GSS 03-JUN-1999
LOCUS CNS008F6/c
DEFINITION Drosophila melanogaster genome survey sequence T73 end of BAC #
BACR17N15 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL052505
VERSION AL052505.1 GI:4933656
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers

```

FEATURES

```
source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="RPC1-98"
/clone="BACR17N15"
/notes="end : TET3"
BASE COUNT      305 a      153 c      198 g      321 t      124 others
ORIGIN

Query Match      6.7%; Score 36.4; DB 82; Length 1101;
Best Local Similarity 39.6%; Pred. No. 2.8;
Matches 80; Conservative 25; Mismatches 97; Indels 0; Gaps 0;

OY 264 aaaaagtgaatgaagccttggcttgaatgacacatgcataaactcaatcaac 323
    ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| :
Db 923 AAAATTTATGKGMMTTTATGTTTTATRAAAMAAAAAATMCCCATTTTCT 864

OY 324 cccattatgcgaaggcaattagcattgcctaaagcctaagtagtcttaccagcg 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 863 GSCATWTTCAATCCCTCYCAAMAAAGTATTATCCCTAAATAATGTTGCAAAAT 804

OY 384 taatgcaactacatacacaagaagcgacaaacacagcctagcagcggtgtgtgt 443
    | : | ||| : | : | ||| : | : | ||| : | : | ||| : | : | |||
Db 803 TCTWAGTTCAMATTAAMAMATGMAAMAAAMATCTGTTGTTWAKGGGTTTCTAT 744

OY 444 tggcttaaacattagcaat 465
    | : | ||| : |||
Db 743 TTKGGTGACACATACYSCTAT 722

RESULT 7
A1138233/c 460 bp mRNA EST 28-OCT-1998
LOCUS      qd76e03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1735420
DEFINITION 3', mRNA sequence.
ACCESSION  A1138233
VERSION    A1138233.1 GI:3644205
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 460)
AUTHORS   NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
JOURNAL   On May 1, 1998 this sequence version replaced gi:2339570.
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CCAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILMIL at:
            www-bio.llnl.gov/bdrrp/image/image.html

Insert Length: 1159 Std Error: 0.00
Seq Primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 451.
Location/Qualifiers
1. .460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:1735420"
/clone="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

```
BASE COUNT      127 a      94 c      96 g      143 t
ORIGIN

Query Match      6.6%; Score 35.8; DB 42; Length 460;
Best Local Similarity 54.1%; Pred. No. 3.3;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 403 aacaaagcgacaaacagcctagcagcggtgtgtgtgtgtgttgaaccattagca 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 AAGACTCCCCAAAACAGTCTGGAAGTCAGAGAGATATGTCCTCAACTACTTGAT 245

OY 463 aatgtggcggttgaagcaagctacactatctatcagaagatgccaatgcaattgtg 522
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 GGCTTACCGATTTTTCAAGAGACATCTATGTCCTATAAAGCAATGCACTGACTAT 185

OY 523 ggcgcctacttgct 537
    || ||| ||| |||
Db 184 CGCTCTCAAGTGTCT 170

RESULT 8
AO784640/c 514 bp DNA GSS 03-AUG-1999
LOCUS      HS_3193_B1 B07.77C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3193 Col-13 Row=D, genomic survey
sequence.
ACCESSION  AO784640
VERSION    AO784640.1 GI:5692264
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 514)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL   Mahairas GG, Wallace JC, Hood L
COMMENT   High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.hsc.washington.edu
            Plate: 3193 Row: D Column: 13
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 514.
            Location/Qualifiers
            1. 514
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate-3193 Col=13 Row=D"
            /clone_id="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
            E-Coli DH10B"
```





Oy	162	ttaatgatttatcagggtttaacttgcacaaatcttgccgtgaagaccgaatttc	221
Db	726	TDTKCKKKTGTGCTKKAADAMWADTWMARARRAAMAAYRGGDKKGSTRDASGHNNHMYTC	667
Oy	222	tgttcaagcagccaagaattaatgatcgaggcgtagtcctgtaaaaaggtgatgtga	281
Db	666	HSNABHMGAHYCNHGCAVYVMNMNRFFAAAAAAATATAAAAWMAADAATAAAAAM	607
Oy	282	tttgtgtcttatgacacatatcgctataactcaactcaataaccattatgcaagg	341
Db	606	NMMBACNMHSYWTCTCACAHMCHMCWCWAHTVCXYATATATAMWAMBACATANAAMR	547
Oy	342	caaatgaagtcatgtgaactaaagtatgctcacggcctaatagcacaataacc	401
Db	546	WRMAWMTVRCDTDADAMATANAKHKNDTWTRWMTADWWANDANRYAADHMDHGH	487
Oy	402	aaacaagaagcagcaaaa	418
Db	486	GWKAMDITWCRHAAMMA	470
RESULT	11		
LOCUS	AM201163		
DEFINITION	AM201163	407 bp mRNA EST 30-NOV-1999	
VERSION	AM201163	se8sf03.y1 Gm-cl027 Glycine max CDNA clone GENOME SYSTEMS CLONE ID	
KEYWORDS	AM201163	Gm-cl027-822 5' similar to TR:080504 080504 F16B2.14 PROTEIN. ;,	
SOURCE	AM201163.1 GI:6481892	EST.	
ORGANISM	soybean.		
REFERENCE	Glycine max		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core		
	eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;		
	Glycine.		
	1 (bases 1 to 407)		
	Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V.,		
	Khanha,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,		
	Beck,C., Wylie,T., Underwood,K., Stepletoe,M., Theising,B., Allen,M.,		
	Bowers,Y., Pearson,B., Swaller,T., Gibbons,W., Pape,D., Harvey,N.,		
	Schurk,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M.,		
	Mccann,R., Waterston,R. and Willson,R.		
	Public Soybean EST Project		
JOURNAL COMMENT	Unpublished (1999)		
	On May 18, 1998 this sequence version replaced gi:3137634.		
	Contact: Shoemaker R/Public Soybean EST Project		
	Public Soybean EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.wustl.edu		
	This clone is available through: Genome Systems, Inc. 4633 World		
	Parkway Circle St. Louis, Missouri 63134 For further information		
	call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)		
	427-3334 or contact: clones@genomesystems.com or		
	info@genomesystems.com web site: www.genomesystems.com		
	High quality sequence stop: 396.		
FEATURES	Location/Qualifiers		
Source	1..407		
	/organism="Glycine max"		
	/db_xref="taxon:3847"		
	/clone="GENOME SYSTEMS CLONE ID: Gm-cl027-822"		
	/clone_1kb="Gm-cl027"		
	/tissue_type="cotyledons of 3- and 7-day-old Williams		
	seedlings"		
	/lab_host="DH10B"		
	/note="Vector: pBluescript II SK+; Site.1: EcoRI, Site.2:		
	XhoI; This cDNA library was constructed from mRNA isolated		
	from cotyledons of 3- and 7-day-old Williams seedlings		
	which were propagated on paper towels with distilled		
	water. The cotyledons were flash-frozen in liquid		

nitrogen, then lyophilized for 72 hours. Unequal amounts of mRNA was used for cDNA synthesis. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-stranded synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemethylated. A modification of Stratagene's first-strand synthesis primer was used. An anchor nucleotide (V=A, C, or G) was added to the 3' end of the primer (GAGAGAGAGAGAGAGAGACATGCTCGAG(T)<sub>18</sub>) to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript(tm) II XR Predigested vector (pBluescript II SK(+)) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 97% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=30). This library was constructed by Dr. Paul Keim and Dr. Virginia Corryell."

Query Match	Similarity	6.4%	Score 34.8	DB 74	Length 407				
Best Local Similarity		54.8%	Pred No. 6.2						
Matches	69	Conservative	0	Mismatches	57; Indels 0; Gaps 0				
OY	171	ttatgcaggtataacttgcaccaaaatttggcgtagaaacccgaatttggtgctcaga							
Dd	224	TAATACAGCTGTTAATTATTTGAGCGCCTATCTTGNGGAGAATAACTGAATCTGTGTGCTGA							
OY	231	cgccaagaatttaatatcatcaggcgtgtgatgccctttaagaagtgtatgaagtcctttggtgc							
Dd	284	GGCTTGAGACAATCAAAAGGCCTGACCAATTCCTAGCATGATCAATGCTTATGAAGT							
OY	291	ttatgg							
Dd	344	GGATTG							
RESULT 12									
LOCUS	AQ176788/c								
DEFINITION	Hs_3213_AL_E04_T7 CTR Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3213 Col-7 Row-1, genomic survey sequence.								
ACCESSION	AQ176788								
VERSION	AQ176788.1								
KEYWORDS	GSS.								
SOURCE	human.								
ORGANISM	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
	Eutheria; Primates; Catarrhini; Homnidae; Homo.								
REFERENCE	1 (bases 1 to 556)								
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.								
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome								
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)								
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University Of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel.: (206) 616-3618								



Db 324 TGCATGAGGNAGATGAGAGAAAGATGAGAAAGATCATGCTTTGAAGAGAAAGACTT 265  
QY 165 tggatattatgcaggtataactttgacccaaatttggcgtagaacccgaattgttgg 224  
Db 264 TGAAGATACCAAGAGAAAGATGCTTGGACACTA-TTTGGCGGAGAAATCTGATCTTCG 206  
QY 225 ttccagcccaaaagaattatgcagcggtgagtcctgtaaaagtgatgtgaagtcctt 284  
Db 205 CTCCTCCCAACAGAGACATGAGAGCTTAAGGAGAGCTTATATCTGATATTAATGAAGCTCTT 146  
QY 285 tggctgtatgacacatacgtctataactc 318  
Db 145 TTCTGATAGAGATCATGATGCAAGATCTGACC 112

RESULT 15  
LOCUS A1794544 598 bp mRNA EST 02-JUL-1999  
DEFINITION fc45f09.y1 zebrafish Washu MPIMG EST Danio rerio CDNA 5', mRNA  
sequence.  
ACCESSION A1794544  
VERSION A1794544.1 GI:5342260  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprinidae; Rasbora; Danio.  
1 (bases 1 to 598)  
Clark, M., Johnson, S.L., Lebrach, H., Lee, R., Li, F., Marra, M.,  
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
Washu zebrafish EST Project 1998  
Unpublished (1998)  
On Jun 5, 1998 this sequence version replaced gi:3189700.  
Other ESTs: fc45f09.x1  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zdrafish@watson.wustl.edu  
CDNA Library Preparation: Matthew Clark, CDNA Library Arrayed by:  
Matthew Clark, DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
Ressourcenzentrum Primardatenbank, Berlin, Germany (web address:  
www.rzpd.de)  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 498.  
Location/Qualifiers  
1..598  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone\_lib="zebrafish Washu MPIMG EST"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="X11-blue MRF"  
/note="Vector: pSPOR1; Site 1: NotI; Site 2: SalI; 1st  
strand cDNA was primed with a Not I - oligo(dT)15 primer  
[5' pGACTAGTTTGAATCGGAGCGGCCCTTTTCTTTTCTTTT3'];  
double-stranded cDNA was ligated to Sal I adaptors (BRL),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lebrach lab; ICRF, London and Max Planck  
Institut fuer Molekulare Genetik Berlin). cDNAs for EST  
analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single cDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."

BASE COUNT 199 a 96 c 84 g 219 t  
ORIGIN

Query Match 6.4%; Score 34.6; DB 60; Length 598;  
Best Local Similarity 49.2%; Pred. No. 8;  
Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 145 aacggtaaagacacgcgtatgtatgtacaggttataacttgacccaaatttggc 204  
Db 161 AACAAATTGAAATATGCAATAGCAAAATGTAATTTTGAAGTTTAAACGGCCAAAATAATC 240  
QY 205 gtgaaccggaattgttggtcagacgcgcaagaattatgacagcggtgagtcctgta 264  
Db 241 AAAAAGCTCAATATCTTCTTTTAAACAAAAGATTTTCTCCCTTGCTTTTCACTCA 300  
QY 265 aaagtgatgtgaagtccttgggtcctatgacacatacgtctataactcaataacc 324  
Db 301 AATGTACAAATTTTGTCAATGTAGTGATGCCAATTAATGCGAATAACGCTTGTTATA 360  
QY 325 ccatt 329  
Db 361 CAGTT 365

Search completed: May 20, 2000, 10:35:50  
Job time: 12797 sec

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2000, 11:13:12 ; Search time 34.13 Seconds

(Without alignments)  
124.919 Million cell updates/sec

Title: US-09-164-714-7

Sequence: 1 MKTKTLTLLAVSASSLLAMSA.....SYNTLSEDAANAISLGAHLAF 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	12.6	188	1	R26415
2	112.5	12.3	361	1	R05999
3	112.5	12.3	361	1	R66877
4	112	12.3	359	1	R66294
5	112	12.3	362	1	P90098
6	108	11.8	188	1	R25674
7	108	11.8	188	1	R70744
8	108	11.8	188	1	W18380
9	106.5	11.7	162	1	R96207
10	106.5	11.7	338	1	R85450
11	106.5	11.7	568	1	R96210
12	100	11.0	288	1	W27720
13	100	11.0	1287	1	R79944
14	97	10.6	211	1	W20755
15	97	10.6	211	1	W55324
16	95.5	10.5	1290	1	W98269
17	95	10.4	1288	1	W55547
18	95	10.4	1288	1	W55685
19	94.5	10.4	330	1	W21678
20	93.5	10.3	257	1	W55462
21	91	10.0	1296	1	R41198
22	89.5	9.8	767	1	W46272
23	89.5	9.8	1267	1	R04232
24	88.5	9.7	174	1	W94460
25	87.5	9.6	443	1	W94460
26	86.5	9.5	363	1	R66878
27	85.5	9.4	711	1	W73025
28	85.5	9.4	758	1	W46270
29	85	9.3	621	1	W55545
30	85	9.3	637	1	W55739
31	85	9.3	691	1	W73027
32	85	9.3	691	1	W73032
33	84.5	9.3	255	1	W55201
34	84.5	9.3	255	1	W55477

35	84	9.2	174	1	W04893	Proteinase K resis
36	83	9.1	759	1	W46271	Moraxella catarrha
37	83	9.1	759	1	W46273	Moraxella catarrha
38	83	9.1	1992	1	W04505	Moraxella 200 kDa
39	82.5	9.0	595	1	W98283	H. pylori GHPO 125
40	80.5	8.8	341	1	R07042	P3 gene product of
41	80.5	8.8	342	1	R66879	H. influenzae prot
42	80.5	8.8	342	1	R66880	H. influenzae prot
43	80.5	8.8	1612	1	W65088	R. prowazekii S-la
44	79.5	8.7	175	1	W04892	Proteinase K resis
45	79.5	8.7	483	1	W73031	Helicobacter pylor

## ALIGNMENTS

RESULT	ID	Score	Match	Length	DB ID	Description
1	R26415	115	12.6	188	1	R26415
AC	R26415	112.5	12.3	361	1	R05999
DT	R26415	112.5	12.3	361	1	R66877
DE	R26415	112	12.3	359	1	R66294
KW	R26415	108	11.8	188	1	R25674
KW	R26415	108	11.8	188	1	R70744
OS	R26415	108	11.8	188	1	W18380
KS	R26415	106.5	11.7	162	1	R96207
FM	R26415	106.5	11.7	338	1	R85450
FT	R26415	106.5	11.7	568	1	R96210
FT	R26415	100	11.0	288	1	W27720
FT	R26415	100	11.0	1287	1	R79944
FT	R26415	97	10.6	211	1	W20755
FT	R26415	97	10.6	211	1	W55324
FT	R26415	95.5	10.5	1290	1	W98269
FT	R26415	95	10.4	1288	1	W55547
FT	R26415	95	10.4	1288	1	W55685
FT	R26415	94.5	10.4	330	1	W21678
FT	R26415	93.5	10.3	257	1	W55462
FT	R26415	91	10.0	1296	1	R41198
FT	R26415	89.5	9.8	767	1	W46272
FT	R26415	89.5	9.8	1267	1	R04232
FT	R26415	88.5	9.7	174	1	W94460
FT	R26415	87.5	9.6	443	1	W94460
FT	R26415	86.5	9.5	363	1	R66878
FT	R26415	85.5	9.4	711	1	W73025
FT	R26415	85.5	9.4	758	1	W46270
FT	R26415	85	9.3	621	1	W55545
FT	R26415	85	9.3	637	1	W55739
FT	R26415	85	9.3	691	1	W73027
FT	R26415	85	9.3	691	1	W73032
FT	R26415	84.5	9.3	255	1	W55201
FT	R26415	84.5	9.3	255	1	W55477

35	84	9.2	174	1	W04893	Proteinase K resis
36	83	9.1	759	1	W46271	Moraxella catarrha
37	83	9.1	759	1	W46273	Moraxella catarrha
38	83	9.1	1992	1	W04505	Moraxella 200 kDa
39	82.5	9.0	595	1	W98283	H. pylori GHPO 125
40	80.5	8.8	341	1	R07042	P3 gene product of
41	80.5	8.8	342	1	R66879	H. influenzae prot
42	80.5	8.8	342	1	R66880	H. influenzae prot
43	80.5	8.8	1612	1	W65088	R. prowazekii S-la
44	79.5	8.7	175	1	W04892	Proteinase K resis
45	79.5	8.7	483	1	W73031	Helicobacter pylor

Query Match 12.6%; Score 115; DB 1; Length 188;  
Best Local Similarity 27.1%; Pred. No. 0.0005;  
Matches 55; Conservative 28; Mismatches 82; Indels 38; Gaps 11;

QY	1	MKTTLTLLAVSASSLLAMSAANAISYGNSADAPYVCAKICGVDAKQINGNTAYGIYAG	60
DB	1	MKTTLTLLAVSASSLLAMSAANAISYGNSADAPYVCAKICGVDAKQINGNTAYGIYAG	60
QY	61	YNFQONFGEAFVGSDA-----KEFNAGVSPV-----KGDVSGFAYGT-----YRY	103
DB	55	YEDDS-----PYSFISLSLYLVDRQASGVPEGIHYHDKFEV-----YGLMNGPARYL	106
QY	104	NEINTPYAKGKLGIAK-TKVDVTSRNATYYSNK--SDKTSIAGGVGVGFRPLANVGVEA	160
DB	107	S-DNFSLYALAGVGTATKFEHSTQDDPSFNKISRKTFANGAGVGMPLNIVVDV	165
QY	161	SY---NTLSEDAANAISLGAHLAF	180

Db	166	GYEGSNISSTKINGFNVCGYRF	188
RESULT	2		
ID	R05999		
AC	R05999 standard; protein; 361 AA.		
DT	04-DEC-1990 (first entry)		
DE	P2 gene product of Haemophilus strain 1H.		
KW	P2 gene; strain 1H; influenza; vaccine; ds.		
OS	Haemophilus influenzae.		
PN	EP-378929-A.		
PD	25-JUL-1990.		
PF	22-DEC-1989; 313573.		
PR	23-DEC-1988; GB-030124.		
PR	01-FEB-1989; GB-002178.		
PA	(CONN)-CONNAUGHT LAB LTD.		
PI	Munson RS, Tolian RW, Chong P, Fahlm R, McVerry P, Klein M;		
DR	WPI; 90-225607/30.		
DR	N-PSDB; 005372.		
PT	Gene coding for protein P2 of Haemophilus influenzae type-B -		
PT	used for developing vaccines for protection against disease		
PT	caused by the organism.		
PS	Claim 1; Page 8; 15pp; English.		
CC	Peptide may be used as a vaccine to the disease caused by HI		
CC	type b, as carrier for conjugation to oligosaccharide derived		
CC	from Haemophilus. Protein may be used with other haptens as		
CC	T-cell dependant antigen and carrier.		
SO	Sequence 361 AA;		

Query Match	12.3%	Score 112.5;	DB 1;	Length 361;
Best Local Similarly	28.1%	Pred. No. 0.002;		
Matches 52;	Conservative 25;	Mismatches 65;	Indels 43;	Gaps 11;

[illegible]

RESULT 3  
ID R66877 standard; Protein; 361 AA.  
AC R66877;  
DT 22-AUG-1995 (first entry)  
DE H. influenzae p2 encoded by pNV-1 CDNA.  
KW Haemophilus influenzae type b; outer membrane protein; P2.  
OS Haemophilus influenzae type b.  
PN WC9503069-A.  
PD 02-FEB-1995.  
PE 22-JUL-1994; U08326.  
PR 23-JUL-1993; U5-096181.  
PA (NAVA-) NORTH AMERICAN VACCINE INC.  
PI Liang S, Pullen JK, Soper TS, Tai JY;  
DR WPI: 95-075024/10.  
DR N-PSDS: 084354.  
PT Production of haemophilus influenzae Hibp2 protein - for the  
PT prevention of bacterial meningitis in animals  
PT Example; Figure 4; 63pp; English.  
CC Total genomic DNA was isolated from H. influenzae type b strain Eagan

CC and used as a template for two  $\phi 2$  specific oligos in a PCR. The 5'  
CC oligo (Q88344) was designed to be 40 bp 5' of the AUG start codon.  
CC The 3' oligo (Q88349) was designed to be 300 bp 3' of the stop codon.  
CC Both oligos contd. SalI restriction sites. The amplified DNA was  
CC used to transform E.coli and the resulting colonies isolated and  
CC analysed. Clones contg. a 1.4 kb fragment were chosen for DNA  
CC sequence analysis. One clone, designated pNV-1, was found to be  
CC identical to the published sequence for Hib strain Mmm A (Munson,  
CC R. and Tolan, R.W., Infection & Immunity 57:88-94, Jan. 1983). The  
CC synthetic oligos used to sequence the DNA are indicated in the FT.  
CC The direction of sequencing was 5'-3'. The rest of pNV-1 is  
CC identical to pUC18. The lac promoter is adjacent to the lower  
CC SalI site.  
SQ Sequence 361 AA;

Query Match	12.3%	Score 112.5;	DB 1;	Length 361;
Best Local Similarity	28.1%;	Pred. No. 0.002;		
Matches 52; Conservative	25;	Mismatches 65;	Indels 43;	Gaps 11;

0y 5 KTLAVASASLLAMSANAIAISYGNADQPPYVGAKIGY-----DAKIQG-----K 51  
Db 3 KTLALILVGAAPASASANAAYVYNNEG-TNVELGGGLSTIAEOSNSTVDNQQOQGALRNQ 61  
0y 52 NTAYGIVAGYNPDQNF-----GVAEFV-----GSDAKEFNAGSVPYKGDVSKSGANGTY 101  
Db 62 GSREHIKATIHNFGEFTYAGCYLETREFVTWKASENGSD-NF-----GDTTS-----KY 106  
0y 102 RYNFNTPEFYPAKGLGIKATYVD-VTSNATYTS--NKSDEKTSIAGG-VGVGFRPLANVG 157  
Db 107 AAVTLGNNAFGEVYKLGRAKTIADGITTSAEDEKCYVLNNSDIPIPSGNIVGITTGFGIDGLV 166  
0y 158 YEAST 162  
Db 167 LGANY 171

RESULT 4  
R66294  
R66294 standard; Protein; 359 AA.  
AC R66294:  
DT 09-AUG-1995 (first entry)  
DE Non-typable Haemophilus influenza (NTHI) fimbria protein.  
KW Fimbria protein; vaccine; otitis media.  
OS Haemophilus influenza strain 1128.  
FH Key  
FT Location/Qualifiers  
FT 22..33  
FT /label= amino terminus  
FT 234..249  
FT /label= internal CNBR fragment  
PD MO9426304-A.  
PD 24-NOV-1994.  
PF 12-MAY-1994; 005477.  
PF 18-MAY-1993; US-065442.  
PI (OHIO-) OHIO STATE RES FOUND.  
PI Bakaletz LO, Kolattukudy PE, Sirakova T;  
DR WPI; 95-006359/01.  
DR N-PSDB; Q78916.  
PT Vaccine comprising non-typable Haemophilus influenza fimbria protein - useful in studying, preventing or reducing the severity of otitis media, also fimbria protein and DNA.  
PS Disclousre: Fig 5; 45pp; English.  
CC The fimbria proteins from 15 randomly selected type b and non-typable clinical isolates of Haemophilus influenzae share common epitopes. Thus fimbria isolated from non-typable Haemophilus influenzae 1128 strain is a particularly suitable immunogen to protect against the different non-typable H<sub>2</sub> influenzae that cause otitis media. Fimbria protein is produced by culturing a transformed microbial host, pref. E.coli. Sporodoptera frugiperda or a mucosal pathogen. Fimbria protein (FP) produced by this process is claimed. The FP protein migrates in polyacrylamide gels to a posn. equiv. to a mol. wt. of 25.5 kD or 37.5 kD.  
SO Sequence 359 AA:

Query Match	12.3%	Score 112:	DB 1:	Length 359;
Best Local Similarity	28.2%	Pred. No. 0.0023;		
Matches 55;	Conservative 18;	Mismatches 80;	Indels 42;	Gaps 11.
QY	5	KTLLVASASLLIAMSANAAISYGNAGADQAPYGA	KIGQ---	VDAKQING-----
DB	3	KTALVLVYAGLAASVADQAAPOENTF----	IAGYAKAGGCSFIDG	INNNGAIKKGLSSNW 58
QY	51	--KNT-AYGIYAGYNF--DQNGVEAEFVGSD--	AKEFNAGVSPVKDVSFGAY---	99
DB	59	GYSRNTFFYGVGCGVQILNDQNFGLAELGYNDE	PERAKRLREGKPKAKH--	TNHGAYLSL 116
QY	99	-GTYRYNTINTPEYAKGKLGIAKTIVD--	VTSRNATYYSNKSDDTSLAG--	-VGVGFK 151
DB	117	KGSYE--VVLGIDLVYGRAGVALVRSDDYK	FYEDANGTDHKKGRHTARASGLFAVGA	EYA 173
QY	152	PLANGVEASNYLS 166		
DB	174	VLPELAVRLKETOWLT 188		
RESULT	5			
P90098				
ID	P90098	standard; protein;	362 AA.	
AC	P90098;			
DT	1-NOV-1989	(first entry)		
DE	P2 antigen of Haemophilus influenzae			
KW	Haemophilus influenzae P2 antigen; vaccines; pathogenic			
KW	type b strains.			
OS	Haemophilus influenzae type b			
PN	EP-320289-A.			
PD	14-JUN-1989.			
PE	09-DEC-1988;	311691.		
PR	10-DEC-1987;	US-249482.		
PA	(TEXA) Univ of Texas Syst.			
P1	Hansen EJ;			
DR	WPI; 89-174562/24.			
PT	DNA encoding Haemophilus influenzae P2 antigen			
PT	and in diagnostic hybridisation assays.			
PS	Disclosure; fig 6; 20pp; English.			
CC	P2 antigen of Haemophilus influenzae (see corresp. N90033).			
CC	Isolated from type b strains, the P2 antigens are used as vaccines			
CC	and to detect P2-specific RNA or DNA. Segments are used as antigens			
CC	include amino acids 21-361, 175-197, 260-275, 296-311, 33-353,			
CC	or their functional equivalents.			
CC	Sequence 362 AA;			

R25674  
 R25674 standard; protein; 188 AA.  
 13-JAN-1993 (first entry)  
 pagc.  
 Salmomella positive regulon; phop; vaccination.  
 Salmomella typhimurium.  
 09-JUL-1992.  
 18-DEC-1991; U09604.  
 18-DEC-1980; US-629602.  
 (GEHO ) GEN HOSPITAL CORP.  
 (HARD ) HARVARD COLLEGE.  
 Mekalanos Jv, Miller St;  
 WPI; 92-250084/30.  
 N-PSDB; Q26694.  
 New anti-salmomella vaccine contg. live attenuated mutants -  
 useful to protect against typhoid fever and related diseases, and  
 for detecting salmomella  
 Disclosure, Page 49-51; 57pp; English.  
 pagc is a Salmomella positive regulon (phop) regulated gene. A  
 mutation at pagc confers a virulence defect. Within the scope of  
 the invention, this gene was mutated, such that it included a  
 virulence attenuating mutation. This mutation does not result in  
 the constitutive expression of a gene under the control of the phop  
 regulatory region. The vector sequence encoding this sequence may  
 be used to transform a cell and then the pagc gene product produced.  
 The mutated pagc gene can be used to produce a vaccine comprising a  
 Salmomella cell which has been attenuated by the above mutation.  
 These vaccines can be administered orally thus avoiding resistance  
 by patients to vaccination by infection.  
 Sequence 188 AA;

Query Match	12.3%;	Score 112;	DB 1;	Length 362;
Best Local Similarity	28.0%;	Pred. No. 0.0023;		
Matches	52;	Conservative 25;	Mismatches 65;	Indels 44;
				Gaps 11;
Oy	5	KTLLAVSASLLAMANAIAISYGSNADAPYVGARIGOV-----DAKOING----	51	
		: : :      :   : : : :   : : : :   : :		
Db	3	KTLLAALIYGAFAPASANAIVYNNEC--INVELGRLISIAECSNNSTYDNOXOQHEALFN	61	
Oy	51	KNTAYGIYAGVNFDFNF-----GVEAEFV-----GSDAKEFNAGVSPYKGVKSFQAYGT	100	
		: : :   :    : : :   : :   : :   : :		
Db	62	QCSRFPIKATHNFGDGFVAGCYLETFRPVTKASENGSD--NF-----GDITS-----K	106	
Oy	101	YRYNININPFYKKGGLGIAKTIKVD--VTSNNATYTS--KMSDKTSLAGG--VGYGFEPLANV	156	
		: : :   :      : :   : :   : :   : :   : :		
Db	107	YAYVLGNKAFGEVRLGKRAKTIADGITSAEDEKEYGLVNLNSDIYPTSGNTVGTGTFKGIIDL	166	
Oy	157	GVEASY	162	
		: :   :		
Db	167	VLGANT	172	
RESULT	6			

```

Query Match          11.8%; Score 108; DB 1; Length 188;
Best Local Similarity 24.2%; Pred. No. 0.0024;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

OY      1 M K T L T L I A V S A S S I L A M S A N A I S Y G N S A D Q P Y V G A K I G C V D A K Q I N G K N T A Y G I A G   60
      || : - - - - - | : | : - - - - - | : | : - - - - - | : | : - - - - - | : | : - - - - - | : | : - - - - -
DB      1 M K N I L L S T L I V T I S Y L V V V N V A Q A D I N A F S V G A R R A Q S K V - - Q D E R K N I G V A V K I - - - - - 54

OY      61 Y N F D N F G V E A E F V G S D A K E F N A G V S P V K G D V K S F G A - - - - - Y G T - - - - - 101
      : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - -
DB      54 - - - - - R Y E D D S P V S F I S S I S Y L X G D Q A G S G S V E P E G I H Y H D K R E V A Y G S L M W P C 102

OY      101 - Y R Y F I N T P P F A K K L G I A K - T K Y D V T S R K A T T Y T S N K - S K T S I A G V G V G F R P L A N V 156
      : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - - : - - - - -
DB      103 A Y R L S - D N F S L A L A G V G V K A T E F E H S T O D D S F S N K I S S R K T G F A M G A G V O M P L E N I 161

OY      157 G V E A S Y - - - N Y L S E D A N A I S L G A H L A F 180
      | : | - - - - - | : | - - - - - | : | - - - - - | : | - - - - - | : | - - - - - | : | - - - - - | : | - - - - -
DB      162 V Y D V G I E G S N I S T K I N G F N G V G I R F 188

RESULT 7
R70744
AC      R70744 standard; Protein; 188 AA.
DT      14-AUG-1995 (first entry)
DE      PacC protein.
KW      Salmonella; vaccine; attenuation; pacC; Phob regulon.
OS      Salmonella typhimurium.
PN      W03502048-A.
PD      19-JAN-1995.
PF      07-JUL-1994; 007658.
PR      09-JUL-1993; US-090526.
PS      06-JUL-1994; US-271354.
PA      (GEHO ) GEN HOSPITAL CORP.
PA      (HARD ) HARVARD COLLEGE.
PI      Mekalanos JF, Miller SI.
DR      WPI; 95-066894/09.
N       N-PSDB; 085099.

```





```
Db 5 VYAAESSISIGYAQSHVKNGETLNDP-----KGFNLKY-----REYELD 46
Qy 66 NEGVAEAEFVGS-DAKEFNAGSPV-KGDYKSGFAGYTRYNF-INTPFYAKKGLGIATK 122
Db 47 NMGVIGSFAYTHQGYDFEYGSNKRFGHGVDDYYSV--TWGSPFRINEYSLVGLGAHAK 104
Qy 123 VDVTSRNATTY--SNKSDKTSIAGVGFGFKPLANVGEASYNLYSEDAHAIS---LGAH 177
Db 105 V-----KASVFDESISAKSTMAYGAGVOFNPLPFVVIDASYETSKLDSIVGTWMLGAG 159
Qy 178 LAF 180
Db 160 YRF 162

RESULT 10
ID R85450 standard; Protein; 338 AA.
AC R85450;
DT 15-FEB-1996 (first entry)
DE Nontypable H. influenzae P5 protein.
KW P5 outer membrane protein; vaccine; otitis media; sinusitis;
KW chronic pulmonary obstructive disease.
OS Haemophilus influenzae.
FH Key Location/Qualifiers
FT misc_difference 195
FT /note= "amino acid at position 195 is not
FT /note= "amino acid at position 311
FT /note= "amino acid at position 311 is not
FT /note= "amino acid at position 311 is not
FT misc_difference 311
FT identified in the specification"
FT EP-680765-A1.
PD 08-NOV-1995.
PR 02-MAY-1995; 302996.
PR 05-MAY-1994; US-210394.
PA (AMCY ) AMERICAN CYANAMID CO.
PI Zlotnick GM:
DR WPI: 95-375029/49.
PT Purified H. influenzae P5 outer membrane protein - used for
PT preventing reducing susceptibility to or treating H. influenzae
PT infections
PS Disclosure: Page 7-8; 16pp: English.
CC Nontypable H. influenzae HI outer membrane protein P5 was isolated
CC by extraction of the outer membrane with detergents and cation-exchange
CC chromatography. P5 (or its peptide fragments) are used in vaccines
CC for prevention of H. influenzae infections implicated in otitis media,
CC sinusitis and chronic pulmonary obstructive disease.
SQ Sequence 338 AA;

Query Match 11.7%; Score 106.5; DB 1; Length 338;
Best Local Similarity 25.6%; Pred. No. 0.0072;
Matches 50; Conservative 18; Mismatches 64; Indels 63; Gaps 11;

Qy 35 YVGAKITG---VDKQING-----KNF-ANGIYGYNF--DQNGVEARE 72
Db 8 YAGKAGGSGFHDGINNGAIKEDSIDTLGYXRNFTFYGVFGGYOILMDNGFLAAE 67
Qy 73 FVGSQ-----AKEFNAGVS-PYKGVKSGFAGYTRYNFINTPFYAKKGLG 117
Db 68 -LGTIDNEFRVFRAGEGKTKAKHTNHGAHLSTLKGSEYVLDGLDY-----GKAG 114
Qy 118 IAKTRVD---VTSRNATTYSNKSDKTSIAGG---VGVGFKPLANVGEASYN---NY 164
Db 115 VALVRSDYKFEAPNSTDAKKGHTARASGLFVAGAEYAVLPBLAVLREYQQLRVCKY 174
Qy 165 LSEDAHAISLGAHIA 179
Db 175 RPODKNAPSINPNTA 189

RESULT 11
R96210
```

```
ID R96210 standard; Protein; 568 AA.
AC R96210;
DT 24-AUG-1996 (first entry)
DE All protein-maltose binding protein fusion protein.
KW Attachment-invasion-locus protein; maltose binding protein;
KW fusion protein; affinity tail; amylose; affinity chromatography;
KW purification; drug delivery; gastrointestinal membrane;
KW transcytosis; bioavailability; enterocyte; Peyer's patch M-cell.
OS Chimeric; Synthetic;.
FH Key Location/Qualifiers
FT region 1..392
FT /note= "Maltose binding protein"
FT region 393..568
FT /note= "Attachment-invasion-locus protein"
FT peptide 425..440
FT /note= "Peptide from extracellular loop-1"
FT peptide 464..482
FT /note= "Peptide from extracellular loop-2"
FT peptide 508..525
FT /note= "Peptide from extracellular loop-3"
FT peptide 551..558
FT /note= "Peptide from extracellular loop-4"

M09613250-A1.
PD 09-MAY-1996.
PR 20-OCT-1995; U13749.
PR 27-OCT-1994; US-331393.
PA (AMGE-) AMGEN INC.
PI Haberfield AD, Jensen-Pippo K;
DR WPI: 96-251447/25.
PT Therapeutic delivery system utilizing bacterial invasion protein - 1s
PT not readily degraded in the gut, enhances systemic bio-availability
PT of therapeutic agents
PS Example 3; Fig 15; 110pp; English.
CC The sequence represents a fusion protein of a Yersinia
CC enterocolitica attachment-invasion-locus (AIL) protein (R96207)
CC with maltose binding protein (MBP, R96208). The MBP N-terminal
CC portion of the fusion protein acts as an affinity tail, allowing
CC efficient recombinant protein purification by amylose affinity
CC chromatography. The receptor binding region of the AIL protein
CC involves all or some of the regions from the 4 extracellular loops.
CC This region retains the binding affinity of the protein, and may be
CC used alone or as part of a fusion protein for drug delivery. The
CC bacterial AIL protein may be complexed with a therapeutic agent to
CC transport the agent across the gastrointestinal membrane barrier by
CC transcytosis to increase bioavailability 5- to 100-fold. The
CC delivery system allows improved transport across enterocytes and
CC Peyer's patch M-cells. The system is not prone to degradation in the
CC gut or early release of biologically active material, and
CC eliminates the need for parenteral administration.
SQ Sequence 568 AA;

Query Match 11.7%; Score 106.5; DB 1; Length 568;
Best Local Similarity 30.1%; Pred. No. 0.014;
Matches 55; Conservative 19; Mismatches 72; Indels 37; Gaps 10;

Qy 10 VASALLMSANAALISY---GNSADQPYGAKIGQVDAKINGKNFYGYAGYNFQD 65
Db 411 VYAAESSISIGYAQSHVKNGETLNDP-----KGFNLKY-----REYELD 452
Qy 66 NEGVAEAEFVGS-DAKEFNAGSPV-KGDYKSGFAGYTRYNF-INTPFYAKKGLGIATK 122
Db 453 NMGVIGSFAYTHQGYDFEYGSNKRFGHGVDDYYSV--TWGSPFRINEYSLVGLGAHAK 510
Qy 123 VDVTSRNATTY--SNKSDKTSIAGVGFGFKPLANVGEASYNLYSEDAHAIS---LGAH 177
Db 511 V-----KASVFDESISAKSTMAYGAGVOFNPLPFVVIDASYETSKLDSIKVGTWMLGAG 565
Qy 178 LAF 180
Db 566 YRF 568
```

```

RESULT 12
W27720
ID W27720 standard; Protein: 288 AA.
AC W27720;
DE 11-MAY-1998 (first entry)
DR H. pylori VacA protein autotransporter region.
KW VacA gene; autotransporter; diagnostic; therapy.
OS Gram-negative bacteria; surface presented polypeptide.
FH Helicobacter pylori.
FT Key Location/Qualifiers
FT Protein 1..288
FT /note="partial protein sequence"
PD MO9135022-A1.
PD 25-SEP-1997.
PF 15-MAR-1996; E01130.
PR 15-MAR-1996; WO-E01130.
PA (PLNC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Jose J. Maurer J., Meyer TF;
PI WPI: 97-480227/44.
DR N-PSDB: T88157.
PT Presentation of peptide(s) on surface of Gram-negative bacteria -
PT via transformation with vector encoding signal peptide, presented
PT peptide and transporter domain of auto-transporter, producing
PT peptide libraries for epitope mapping
PS Claim 8; Fig 24; 84pp; German.
CC This sequence represents encodes an autotransporter membrane integration
CC region from the H. pylori VacA gene. This region is involved in a novel
CC method which allows the presentation of stable fusion polypeptides on the
CC surface of Gram-negative bacteria which can be released into the
CC surrounding media. The method can be used to produce a variegated
CC population of surface-presented polypeptides, so that bacteria expressing
CC polypeptides with particular properties can be identified and
CC simultaneously selected, e.g. for epitope mapping or selection of ligands
CC with the highest affinity for antibodies, major histocompatibility
CC complex (MHC) molecules or other components of the immune system.
CC Selected polypeptides can be used diagnostically, e.g. to screen sera or
CC antibody banks, and polypeptide expressing cells may be used as live
CC vaccines. They may be used therapeutically, e.g. when the polypeptide is
CC an antibody, to remove or concentrate pollutants, inactivate toxins,
CC prepare and process food, prepare washing compositions and label cells.
CC Selected bacteria can be stored, reproduced and replicated on a large
CC scale as individual clones.
SQ Sequence 288 AA:

```

```

Query Match 11.0%; Score 100; DB 1; Length 288;
Best Local Similarity 22.7%; Pred. No. 0.025; 77; Indels 52; Gaps 10;
Matches 48; Conservative 34; Mismatches 77;
QY 9 AVSASSILAMSANAISYGNADAPYVGAKI-----GOVDAKQINGKN 52
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26 AIGGTSLNS-GGNASL-YETGSAGVDAYLNGEYEAIVGSGSYSSFNQANLNGANN 83
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 53 TAYGIAGYNFDON-FGVEAE-FVGSDAKEFNAGVSPVKDYSF-----GAYGT 100
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 84 TNGGVYSRIEFAHQHEPDEAOGALGSDOSLNFKSLALRDLNOSYNYLAYSAATRASYG- 143
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 YARYNFTPEFYAKGKGIKTKYDVTSRNATITYSNKSDKTSLAGVGVGKPLANGVFA 160
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 143 YDPAEPRNALVLEKPSVGYSTNHLGSTNFK-----SNSNOKVALKNAGASSOHLFNASANVEA 198
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 SYNY-----LSE-----DANAISL 174
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 199 RYYTGDTSYFYMNAGVLOEFANFGSSNAVSL 229
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 13
ID R79944 standard; Protein: 1287 AA.
AC R79944;
DE 26-MAR-1996 (first entry)
DE Helicobacter pylori vacuolating toxin.

```

```

KW Vacuolating toxin; vaccine; immunisation; therapy; mutant;
KW infection; Helicobacter pylori.
OS Helicobacter pylori.
PN MO9522988-A1.
PD 31-AUG-1995.
PF 23-FEB-1995; U02219.
PR 23-FEB-1994; US-200232.
PA (UYVA-) UNIV VANDERBILT.
PI Blaser MJ, Cover TL;
PI WPI: 95-311383/40.
DR N-PSDB: T04132.
PT Isolated DNA encoding Helicobacter pylori vacuolating toxin - useful
PT for immunisation against H. pylori infection
PS Claim 7; Page 43-47; 64pp; English.
CC The nucleic acid encoding the Helicobacter pylori vacuolating toxin
CC and a genetically altered mutant strain of H. pylori which contains
CC a foreign nucleic acid and does not express a functional vacuolating
CC toxin may be used to immunise a subject against H. pylori infection.
CC They may possibly also be used therapeutically.
SQ Sequence 1287 AA;

```

```

Query Match 11.0%; Score 100; DB 1; Length 1287;
Best Local Similarity 24.8%; Pred. No. 0.18;
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;
QY 9 AVSASSILAMSANAISYGNADAPYVGAKI-----GOVDAKQIN----- 50
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1025 AIGGAS-LNNGGNASL-YETGSAGVDAYLN---GQVAYIVGSGSYSSFNQANLNSG 1079
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 50 GKNTAYGIAGYNFDON-FGVEAE-FVGSDAKEFNAGVSPVKDYSF-----GA 97
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1080 ANNTNGVYSRIEFAHQHEPDEAOGALGSDOSLNFKSLALRDLNOSYNYLAYSAATRAS 1139
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 98 YGYRRNFTNTPFYAKGKGIKTKYDVTSRNATITYSNKSDKTSLAGVGVGKPLANGV 157
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1140 YG-YDPAEPRNALVLEKPSVGYSTNHLGSTNFK-----SNSNTKVALNSGSSOHLFNASAN 1194
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 158 VEASVNY-----LSE-----DANAISL 174
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1195 VEARYYTGDTSYFYMNAGVLOEFANFGSSNAVSL 1228
  | : | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
W20755
ID W20755 standard; Protein: 211 AA.
AC W20755;
DE 15-JUL-1997 (first entry)
DE H. pylori protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
KW outer membrane; cell envelope; transporter.
OS Helicobacter pylori.
PN MO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTRA) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
PI WPI: 97-052306/05.
DR N-PSDB: T68008.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Disclosure; Pages 1168-1169; 1481pp; English.
CC The present sequence is a Helicobacter pylori protein of unknown
CC function. The protein may be used in a vaccine to prevent or treat
CC H. pylori infection or to identify H. pylori polypeptide binding
CC compounds, useful as potential H. pylori life cycle activators or
CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
CC determined from overlapping contigs generated by mechanically shearing

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 20, 2000, 12:18:21 ; Search time 27.94 Seconds

(Without alignments)

93.180 Million cell updates/sec

Title: US-09-164-714-7

Perfect score: 912  
Sequence: 1 MTKLTLLAVASASSLLAMSA.....SYNYLSEDAANLSLGHHLAF 180Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/PCRTUS.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113	12.4	359	1 US-08-457-997B-2	Sequence 2, Appl1
2	112.5	12.3	361	4 PCT-US94-08326-8	Sequence 8, Appl1
3	108	11.8	188	1 US-08-286-767-3	Sequence 3, Appl1
4	106.5	11.7	566	4 PCT-US95-13749-5	Sequence 5, Appl1
5	105.5	11.6	338	1 US-08-210-394-1	Sequence 1, Appl1
6	103	11.3	188	1 US-08-466-715-5	Sequence 5, Appl1
7	103	11.3	188	1 US-08-486-719-5	Sequence 5, Appl1
8	103	11.3	188	1 US-08-476-100-5	Sequence 5, Appl1
9	103	11.3	188	3 US-08-475-749-5	Sequence 5, Appl1
10	103	11.3	188	3 US-08-475-749-5	Sequence 5, Appl1
11	100	11.0	1287	1 US-08-200-232-2	Sequence 2, Appl1
12	100	11.0	1287	4 PCT-US95-02219-2	Sequence 2, Appl1
13	100	11.0	1287	4 PCT-US95-02219A-2	Sequence 2, Appl1
14	87.5	9.6	363	4 PCT-US94-08326-10	Sequence 10, Appl1
15	83.5	9.2	455	2 US-08-472-172-4	Sequence 4, Appl1
16	83	9.1	759	2 US-08-450-351-2	Sequence 2, Appl1
17	83	9.1	759	2 US-08-450-351-4	Sequence 4, Appl1
18	80.5	8.8	342	4 PCT-US94-08326-12	Sequence 12, Appl1
19	80.5	8.8	342	4 PCT-US94-08326-14	Sequence 14, Appl1
20	80.5	8.8	1612	1 US-08-169-927-2	Sequence 2, Appl1
21	75.5	8.3	663	1 US-08-765-081-7	Sequence 7, Appl1
22	74.5	8.2	792	3 US-08-433-522A-8	Sequence 8, Appl1
23	74.5	8.2	1394	4 PCT-US95-10661A-2	Sequence 2, Appl1
24	74	8.1	703	4 PCT-US95-06994-9	Sequence 9, Appl1
25	73.5	8.1	340	2 US-08-355-844-1	Sequence 1, Appl1
26	73.5	8.1	340	4 PCT-US95-16126-1	Sequence 1, Appl1
27	72.5	7.9	205	1 US-08-277-231A-5	Sequence 5, Appl1
28	72.5	7.9	205	2 US-08-473-750-8	Sequence 8, Appl1
29	72.5	7.9	205	2 US-08-477-326-8	Sequence 8, Appl1

30	72	7.9	251	1 US-08-209-747-8	Sequence 8, Appl1
31	72	7.9	251	1 US-08-458-298-8	Sequence 8, Appl1
32	72	7.9	459	1 US-08-472-172-2	Sequence 2, Appl1
33	71.5	7.8	1912	1 US-08-409-995-4	Sequence 4, Appl1
34	71	7.8	584	1 US-08-313-288B-17	Sequence 17, Appl1
35	71	7.8	593	1 US-07-961-522-4	Sequence 4, Appl1
36	71	7.8	593	1 US-08-217-438-4	Sequence 4, Appl1
37	71	7.8	593	1 US-08-217-438-5	Sequence 5, Appl1
38	71	7.8	593	1 US-08-487-890A-100	Sequence 100, App
39	71	7.8	593	1 US-08-321-978-4	Sequence 4, Appl1
40	71	7.8	593	2 US-08-710-584-4	Sequence 100, App
41	71	7.8	593	2 US-08-478-435-100	Sequence 100, App
42	71	7.8	593	2 US-08-337-483-100	Sequence 100, App
43	71	7.8	593	2 US-08-478-373-100	Sequence 100, App
44	71	7.8	593	3 US-08-474-671-100	Sequence 100, App
45	71	7.8	593	3 US-08-474-671-100	Sequence 100, App

## ALIGNMENTS

```
RESULT 1
US-08-457-997B-2
: Sequence 2, Application US/08457997B
: Patent No. 5766608
:
: GENERAL INFORMATION:
: APPLICANT: Kolattukudy, P. E.
: TITLE OF INVENTION: Otitis Media Vaccine
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: Calfee, Halter and Griswold
: STREET: Suite 1800 800 Superior Avenue
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: U.S.A.
: ZIP: 44114-2688
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457, 997B
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Goltick, Mary E.
: REGISTRATION NUMBER: 34,829
: REFERENCE/DOCKET NUMBER: 22727/00102
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (216) 622-8458
: TELEFAX: (216) 241-0816
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 359 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-457-997B-2
```

Query Match 12.4%; Score 113; DB 1; Length 359;  
Best Local Similarity 28.2%; Pred. No. 0.00011;  
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;

```
OY 5 KTLVAVSASSLLAMANAISYGNADAPYGCATGQ---VDAQING----- 51
   ||:::| | | | | | | | | | | | | | | | | | | | | |
DB 3 KTAIALVAVAGLAASVADAPQENTF---YAGVRAAGGSGFDGJNNNGAIKGLSSNY 58
   ||:::| | | | | | | | | | | | | | | | | | | | | |
OY 51 ---KMT-AYGIYAGNF---DONFGVEAEFVGS---AKEFNAGVSPVKGDVKSGFAY--- 99
   ||:::| | | | | | | | | | | | | | | | | | | | | |
DB 59 GYRRNTFTYVAGVGQIILNQDFGLAALGIDYDFGRAKLRGKPKAKH--TNGHAYLST 116
   ||:::| | | | | | | | | | | | | | | | | | | | | |
```

```
QY 99 -GTRNYNFINPPFAKRLGLAKTKVD---VTSNMTATYSNKSDKTSLAG---VCYGFK 151
Db 117 KGSE---VLDDGLDYTRKAGVALVRSYKFEYEDANGTRDHKKGRHTRARASGLFAVGAEYA 173
QY 152 PLANVGEASTNYLUS 166
      | : : : :
174 VLPETLAIVRLLEYOMLT 188
      | : : : :
18b
```

## RESULT 2

```

PCT-US94-08326-8
: Sequence 8, Application PC/TUS9408326
: GENERAL INFORMATION:
: APPLICANT: North American Vaccine, Inc.
: APPLICANT: 12103 Indian Creek Court
: APPLICANT: Beltsville, MD 20705
: APPLICANT: Pullen, Jeffrey K.
: APPLICANT: Soper, Thomas S.
: APPLICANT: Liang, Shu-Mei
: TITLE OF INVENTION: A Method For The High Level
: TITLE OF INVENTION: Expression,
: TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
: TITLE OF INVENTION: Protein
: TITLE OF INVENTION: p2 From Haemophilus Influenzae Type b
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/08326
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/096,181
: FILING DATE: 23-JULY-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REFERENCE/DOCKET NUMBER: 1438.001PC01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 361 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US94-08326-8

```

Query Match	12.38	Score 112.5	DB 4	Length 361
Best Local Similarity	28.1%	Pred. No. 0.00013		
Matches 52; Conservative	25	Mismatches 65	Indels 43	Gaps 11

[illegible]

Db	Qy	Db
107	158	167
AVYLGNNKAFGEVRLGRAKTIADGITSAEDEK	VEAST 162	LGANT 171
	: :	

## RESULT

```

US-08-286-767-3
Sequence 3, Application US/08286767
Patent No. 5733760
GENERAL INFORMATION:
APPLICANT: Lu, Yichen
APPLICANT: Miller, Samuel I.
APPLICANT: Killeen, Kevin
TITLE OF INVENTION: SALMONELLA VECTORS ENCODING TRUNCATED
TITLE OF INVENTION: PACG FUSION PROTEIN, METHOD OF MAKING, AND USES THEREO
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: RONALD I. EISENSTEIN, DIKE, BRONSTEIN,
ADDRESSEE: ROBERTS & CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,767
FILING DATE: 05-AUG-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 44420
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6400
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-286-767-3

```

Query Match	11.8%	Score 108	DB 1	Length 188
Best Local Similarity	24.2%	Pred. No. 0.00017		
Matches 50; Conservative	29;	Mismatches 82;	Indels 46;	Gaps 8;

Qy	1	MKIKLTLAASASSLLMSANAALISTYGSNADAPGYAAKIGQVDAKIRINKRNTAYGIYAG	60
	1		
Db	1	MKNITLSTLVTTTSTVLVYVNAQAADITNAFISVGARYAASQSV- ODFKNIRGVNWKY	54
	1		
Qy	61	YNFDQNFGEVEAFEGSDAKFEFNAGVSPVKGVKSCFGA- YGT	101
	61		
Db	54	-----RYEDDSFVSTPISLSLITLYIDRQASGVSEBEGITHYHDKFEVYKGSILWGP	102
	54		
Qy	101	-YRYNFINTPEYAKGKLGIAK-TRVDYTSRNATYTSNK--SDKTSLAGVGEFFKPLANY	156
	101		
Db	103	AYRLIS-DNFSLYALAGVGYVATKREHSTQDGSFSNKKISRKRTGFMAGQGYMNPLENI	161
	103		
Qy	157	GVEASY--NYLSEDAANALISICGAHLAF	180
	157		
Db	162	VVDVGEESNISTSTRKINGEYFNNGYFRE	188
	162		



APPLICATION NUMBER: 07/629,602  
FILING DATE: 18-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/192005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 188 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-486-715-5

Query Match 11.3%; Score 103; DB 1; Length 188;  
Best Local Similarity 27.0%; Pred. No. 0.00059;  
Matches 53; Conservative 28; Mismatches 91; Indels 24; Gaps 10;

QY 1 MKTLKTLAVSASSLAMSANAAISYGSNADAPYVGAKIGVDAKOINCKNTAYGIYAG 60  
DB 1 MKNIIISTLVITTSVLVNVNAQDTNAFSVGYARYAQSKV--ODFNKIRGVNWKY----R 54  
QY 61 YNFDQ--NFGVEAEFYGSDAKFEFNAGVSPV-----KGDK--SFGAYGTYRYNFTNTPF 110  
DB 55 YEDDSVPSFSSISLYLGD-RQASGSVEPEGIYHDKFEYKXGSLWGPARYLS-DNFSL 112  
QY 111 YAKGKGLIAK-TKVDYTSRNATYTSNK--SDKTSLAGVGVGFKPLANVGVEASY---NY 164  
DB 113 YALAGVGYATPKKEHSTOGDSFSNKSISRTKGFAMGAGVQMNPLENIVVDVGYEGSNI 172  
QY 165 LSEDAVAISGAHLAF 180  
DB 173 SSTKINGFNVGGYRF 188

RESULT 7  
US-08-486-719-5  
Sequence 5, Application US/08486719  
Patent No. 5674736  
GENERAL INFORMATION:  
APPLICANT: Miller, Samuel I.  
TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P. C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,719  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/090,526  
FILING DATE: 09-JUL-1993  
PRIOR APPLICATION NUMBER:  
APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
FILING DATE: 18-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/192002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 188 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-486-719-5

Query Match 11.3%; Score 103; DB 1; Length 188;  
Best Local Similarity 27.0%; Pred. No. 0.00059;  
Matches 53; Conservative 28; Mismatches 91; Indels 24; Gaps 10;

QY 1 MKTLKTLAVSASSLAMSANAAISYGSNADAPYVGAKIGVDAKOINCKNTAYGIYAG 60  
DB 1 MKNIIISTLVITTSVLVNVNAQDTNAFSVGYARYAQSKV--ODFNKIRGVNWKY----R 54  
QY 61 YNFDQ--NFGVEAEFYGSDAKFEFNAGVSPV-----KGDK--SFGAYGTYRYNFTNTPF 110  
DB 55 YEDDSVPSFSSISLYLGD-RQASGSVEPEGIYHDKFEYKXGSLWGPARYLS-DNFSL 112  
QY 111 YAKGKGLIAK-TKVDYTSRNATYTSNK--SDKTSLAGVGVGFKPLANVGVEASY---NY 164  
DB 113 YALAGVGYATPKKEHSTOGDSFSNKSISRTKGFAMGAGVQMNPLENIVVDVGYEGSNI 172  
QY 165 LSEDAVAISGAHLAF 180  
DB 173 SSTKINGFNVGGYRF 188

RESULT 8  
US-08-476-100-5  
Sequence 5, Application US/08476100  
Patent No. 5731196  
GENERAL INFORMATION:  
APPLICANT: Miller, Samuel I., III  
APPLICANT: Mekalanos, John J.  
TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P. C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,100  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/090,526  
FILING DATE: 09-JUL-1993  
PRIOR APPLICATION NUMBER:  
APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
FILING DATE: 18-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/192003  
TELECOMMUNICATION INFORMATION:





;; TYPE: amino acid  
;; STRANDEDNESS: not relevant  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-475-749-5

Query Match 11.3%; Score 103; DB 3; Length 188;  
Best Local Similarity 27.0%; Pred. No. 0.00059;  
Matches 53; Conservative 28; Mismatches 91; Indels 24; Gaps 10;

QY 1 MKTTLTLLAVSSSLIAMSANNAISYGNSSADAPYVGAKIGQVDAKQINGRMTAIGIYAG 60  
DB 1 MKNITLSTLVITTSVLVNVNAQADTNAPSVGARYAQSKV--QDFKNIRGVNVKTY---R 54  
QY 61 YNFDO--NFGVEAEFVGSDAKEFNAGVSPV-----KGDVK--SEFAGTYRYNFINTPEF 110  
DB 55 YEDDSPVPSPFISLSLYLD-RQASGSVEPEGIHYHDKFEVKKGSLMVGATYALS-DNPSL 112  
QY 111 YAKGLGIJAK-TKVDVTSRNATTYSNK--SDKTSIAGVGVGFKPLANVGVEAST--NY 164  
DB 113 YALAGVGVTKAFKFEKSHSYDDGSFNSKTSISRKTFAMGAGVQMNPLENTIVDVYEGSNI 172  
QY 165 LSEDAANNAISIGHALAF 180  
DB 173 STKINGFNVGVGRF 188

## RESULT 11

US-08-200-232-2  
Sequence 2, Application US/08200232  
Patent No. 5721349

;; GENERAL INFORMATION:  
;; APPLICANT: Cover, Timothy L.  
;; APPLICANT: Blaser, Martin J.  
;; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
;; TITLE OF INVENTION: AND RELATED METHODS  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESSES:  
;; ADDRESSEE: NEEDLE & ROSENBERG P.C.  
;; STREET: 127 Peachtree Street, Suite 1200  
;; CITY: Atlanta  
;; STATE: Georgia  
;; COUNTRY: USA

;; ZIP: 30303  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/200.232  
;; FILING DATE:

;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Spiralt, Gwendolyn D.  
;; REGISTRATION NUMBER: 36,016  
;; REFERENCE/DOCKET NUMBER: 2200.023  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 404/688-0770  
;; TELEFAX: 404/688-9880

;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1287 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-200-232-2

Query Match 11.0%; Score 100; DB 1; Length 1287;  
Best Local Similarity 24.8%; Pred. No. 0.018;  
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;

QY 9 AVSASSLLAMSAANNAISYGNSSADAPYVGAKIGQVDA-----KQIN----- 50  
DB 1025 AIGGAS-LNNGGNASL-YETSGAYDALN---GQVEAIVGGFGSYGSFNNQANSLNSG 1079  
QY 50 GKNTAYGIYAGYNFQDN-FGVBAE-FVGSDAKEFNAGVSPVKGDKSF-----GA 97  
DB 1080 ANNTNFGVYSRIEANOHEPDAOGALGSDOSLNFKSALLDLQSYNYLAYSAATRAS 1139  
QY 98 YGTYYRNFINTPEFYAKGKGIKTKVDVTSRNATTYSNKSDDKTSIAGVGVGFKPLANYG 157  
DB 1140 YG-YDPAFERNALVLRKPSGVSYNHLGISTNEK----SNSTNRVALNSGSSOHLFNASAN 1194  
QY 158 VEASYN-----LSE-----DANAISL 174  
DB 1195 YEARYYGDTSYFYMNAGVLQEFANRGSNAVSL 1228

## RESULT 12

PCT-US95-02219-2  
Sequence 2, Application PC/TUS9502219

;; GENERAL INFORMATION:  
;; APPLICANT: Cover, Timothy L.  
;; APPLICANT: Blaser, Martin J.  
;; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
;; TITLE OF INVENTION: AND RELATED METHODS  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESSES:  
;; ADDRESSEE: NEEDLE & ROSENBERG P.C.  
;; STREET: 127 Peachtree Street, Suite 1200  
;; CITY: Atlanta  
;; STATE: Georgia  
;; COUNTRY: USA

;; ZIP: 30303  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/02219  
;; FILING DATE:

;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Spiralt, Gwendolyn D.  
;; REGISTRATION NUMBER: 36,016  
;; REFERENCE/DOCKET NUMBER: 2200.023  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 404/688-0770  
;; TELEFAX: 404/688-9880

;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1287 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US95-02219-2

Query Match 11.0%; Score 100; DB 4; Length 1287;  
Best Local Similarity 24.8%; Pred. No. 0.018;  
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;

QY 9 AVSASSLLAMSAANNAISYGNSSADAPYVGAKIGQVDA-----KQIN----- 50  
DB 1025 AIGGAS-LNNGGNASL-YETSGAYDALN---GQVEAIVGGFGSYGSFNNQANSLNSG 1079  
QY 50 GKNTAYGIYAGYNFQDN-FGVBAE-FVGSDAKEFNAGVSPVKGDKSF-----GA 97  
DB 1080 ANNTNFGVYSRIEANOHEPDAOGALGSDOSLNFKSALLDLQSYNYLAYSAATRAS 1139  
QY 98 YGTYYRNFINTPEFYAKGKGIKTKVDVTSRNATTYSNKSDDKTSIAGVGVGFKPLANYG 157

Db 1140 YG-YDFAFFRNALVKPSGVSYNHLGSTNFK-----SNSTNKVALNSGSSQHLFNASAN 1194  
QY 158 VEASYN-----LSE-----DANAISL 174  
Db 1195 VEARYYGDTSFYFMAGVLOEFANFGSSNAVSL 1228

## RESULT 13

PCT-US95-02219A-2  
; Sequence 2, Application PC/TUS9502219A  
; GENERAL INFORMATION:  
; APPLICANT: Cover, Timothy L.  
; APPLICANT: Tumuru, Murat KR  
; APPLICANT: Cao, Ping  
; APPLICANT: Thompson, Stuart A.  
; APPLICANT: Blaser, Martin J.  
; TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
; TITLE OF INVENTION: AND THE RELATED METHODS  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG P.C.  
; STREET: 127 Peachtree Street, Suite 1200  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/02219A  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spratt, Gwendolyn D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 2200.023  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404/688-0770  
; TELEFAX: 404/688-9880  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1287 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-02219A-2

Query Match 11.0%; Score 100; DB 4; Length 1287;  
Best Local Similarity 24.8%; Pred. No. 0.018;  
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;

QY 9 AVASSSLANSANAISYGSADAQPYVGAKIGQYDA-----KQIN----- 50  
Db 1025 AIGAS-LNNGGNSL-YGTSAGVDAYLN--GQVEAIVGFGSYGSSFNQNSLNSG 1079  
QY 50 GKNTAYCIVAGYNDON-FGVEAE-FVSGDAKEFNAGVSPYKGVKSP-----GA 97  
Db 1080 ANNTNFVYSRIFANQHEFDEFAOGALGSDSSLNFKSALIRLNDQSTNYLAYSAAIRAS 1139  
QY 98 YGTYRYNFINTPYAKGKLGIATKTVDTSRNATYTSNKSDDKTSLAGGVGFKPLANVG 157  
Db 1140 YG-YDFAFFRNALVKPSGVSYNHLGSTNFK-----SNSTNKVALNSGSSQHLFNASAN 1194  
QY 158 VEASYN-----LSE-----DANAISL 174  
Db 1195 VEARYYGDTSFYFMAGVLOEFANFGSSNAVSL 1228

RESULT 14

PCT-US94-08326-10  
; Sequence 10, Application PC/TUS9408326  
; GENERAL INFORMATION:  
; APPLICANT: North American Vaccine, Inc.  
; APPLICANT: 12103 Indian Creek Court  
; APPLICANT: Beltsville, MD 20705  
; APPLICANT: Pullen, Jeffrey K.  
; APPLICANT: Soper, Thomas S.  
; APPLICANT: Liang, Shu-Wei  
; TITLE OF INVENTION: A Method For the High Level  
; TITLE OF INVENTION: Expression,  
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane  
; TITLE OF INVENTION: Protein  
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/08326  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/096,181  
; FILING DATE: 23-JULY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REFERENCE/DOCKET NUMBER: 1438.001PC01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 363 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US94-08326-10

Query Match 9.6%; Score 87.5; DB 4; Length 363;  
Best Local Similarity 26.7%; Pred. No. 0.074;  
Matches 47; Conservative 24; Mismatches 62; Indels 43; Gaps 11;

QY 14 SLMSANAAISYGSADAQPYVGAKIGQY-----DKQING-----KNTAYIGAG 60  
Db 14 SLPVSSDPAYVYNNEG-TNVELGRLSLIEQSNSTYDNRQOQALRNGSSRHAT 72  
QY 61 YNFDONF-----GVEAEFY-----GSDAKEFNAGVSPYKGVKSPGAYGTYRYNFINTPF 110  
Db 73 HNFQGGFQAQYGLERFRFTKASENGSD--NF-----GDITS-----KAYVYLGKKA 117  
QY 111 YAKGKLGIATKTV-DVISRNATYTS--NKSDDTSLAGG-VGVGFKPLANVGVEASY 162  
Db 118 FGEVAKLGRAKTIADGITSADKEYGVNLNNSDIPTSGWTGVYTFKIGIDGLVIGANY 173

RESULT 15  
US-08-472-172-4  
; Sequence 4, Application US/08472172  
; Patent No. 5985288  
; GENERAL INFORMATION:  
; APPLICANT: Munson, Jr., Robert S  
; APPLICANT: Grass, Susan



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 20, 2000, 12:19:13 ; Search time 43.2 Seconds

(without alignments) 244.291 Million cell updates/sec

Title: US-09-164-714-7

Sequence: 1 MKTLTILAVASASSILAMSA.....SYNYISEDANAISLAHLAF 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	13.5	178	2 F64124	opacity protein ho
2	123	13.5	178	2 A35123	all protein precur
3	112.5	12.3	359	2 D64050	major outer membra
4	112.5	12.3	361	2 A30542	major outer membra
5	109.5	12.0	264	2 154668	heat resistant agg
6	108	11.8	188	2 A39185	virulence protein
7	108	11.8	365	1 MMBP2	outer membrane por
8	108	11.3	365	1 MMECNC	outer membrane por
9	103	11.3	371	2 S68072	major outer membra
10	103	11.3	371	2 S68069	major outer membra
11	102.5	11.2	213	2 A56152	major 25k outer me
12	101	11.1	386	2 S68062	major outer membra
13	101	11.1	386	2 S68063	major outer membra
14	101	11.1	386	2 S68064	major outer membra
15	100	11.0	1287	2 B53739	vacuolating cyto
16	100	11.0	1291	2 S44983	vacuolating cyto
17	99.5	10.9	172	2 A39189	outer membrane pro
18	98.5	10.8	215	2 JC4087	accessory coloniza
19	96	10.5	385	2 S68066	major outer membra
20	96	10.5	385	2 S68067	major outer membra
21	96	10.5	385	2 S68068	major outer membra
22	96	10.5	385	2 S68070	major outer membra
23	96	10.5	385	2 S68071	major outer membra
24	95.5	10.5	1290	2 G64630	vacuolating cyto
25	95	10.4	211	2 S47347	outer membrane pro
26	95	10.4	575	2 S41310	flagellin A - Camp
27	95	10.4	575	2 I40615	flagellin protein
28	95	10.4	1288	2 E71884	vacuolating cyto
29	90	9.9	185	2 A43309	outer membrane pro
30	90	9.9	196	2 S70957	otnd protein - Vib

31	90	9.9	575	2 S41311	flagellin B - Camp
32	89.5	9.8	1585	2 B69948	phage-related prot
33	89	9.8	171	2 I35173	outer membrane pro
34	89	9.8	359	2 S07298	outer membrane pro
35	88	9.6	346	1 MMECA	outer membrane pro
36	86	9.4	394	2 S34263	outer membrane por
37	86	9.4	623	2 164034	hypothetical prote
38	86	9.4	651	2 E71904	probable outer mem
39	85.5	9.4	356	2 A55016	lysosomal membrane
40	85.5	9.4	483	2 A71958	outer membrane pro
41	85.5	9.4	711	2 A64523	outer membrane pro
42	85.5	9.4	2340	2 B71704	cell surface antiq
43	85	9.3	503	2 T35053	probable solute-bi
44	85	9.3	638	2 B71904	probable outer mem
45	85	9.3	691	2 C64548	outer membrane pro

## ALIGNMENTS

RESULT 1  
F64124  
opacity protein homolog H11457 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: F64124  
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Welman  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: F64124  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1178 <TIGR>  
A:Cross-references: GB:U32823; GB:L42023; NID:g1574281; PIDN:AAC23104.1; PID:g1574294

Query Match 13.5% Score 123.5; DB 2; Length 178;  
Best Local Similarity 27.5%; Pred. No. 0.0014;  
Matches 52; Conservative 30; Mismatches 82; Indels 25; Gaps 8;  
QY 4 LKTLAVASASSILAMSAANAISYGSADAPYGAIG--QYDAKQINGKNT---AYGIIY 58  
Db 1 MKKLLIVMLFTIAL-----SAQAQWYVGDLGASKRIDITHVSSNSPSTQRTS 50  
QY 59 AGYNFDONFGVEAEFVGSDAKEFN-AGVSPYKGDVKSFGANGTYRYNFINPFIYAKKIG 117  
Db 51 VGAYADKNFRLAVDYNTNGKYVANYADVDVSLKGSIGLITGFYDFDLADFPYV--GVR 108  
QY 118 IAKTVVDYTSNATYTSKSKDKTSLAGV---GYGFRLAVNGVNAS--VYIASEDNA 171  
Db 109 VSTNGADVTA-NARYRYIEAFETTRIGISGALAGVAKTDVNAINTIEYRLASNVSD 167  
QY 172 ISGAHLAF 180  
Db 168 VGKAGKGLRF 176  
RESULT 2  
A35123  
all protein precursor - Yersinia enterocolitica  
C:Species: Yersinia enterocolitica  
C:Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 08-Oct-1999  
C:Accession: A35123  
R:Miller, V.L.; Bliska, J.B.; Falkow, S.  
J. Bacteriol. 172, 1062-1069, 1990  
A:Title: Nucleotide sequence of the Yersinia enterocolitica All gene and characterization  
A:Reference number: A35123; MUID:90130261  
A:Accession: A35123  
A:Status: preliminary



heat resistant agglutinin 1 - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 08-Oct-1999  
C:Accession: I54668; 169133; 169134  
R:Lutwyche, P.; Rupp, R.; Cavanagh, J.; Warren, R.A.; Brooks, D.E.  
Infect. Immun. 62, 5020-5026, 1994  
A:Title: Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglut  
A:Reference number: I54668; MUID:95012721  
A:Accession: I54668  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-264 <RES>  
A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AC13751.1; PID:g463911  
A:Accession: I69133  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 15-264 <RE2>  
A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AC13752.1; PID:g463912  
A:Accession: I69134  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 18-264 <RE3>  
A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AC13753.1; PID:g463913

Query Match 12.0%; Score 109.5; DB 2; Length 264;  
Best Local Similarity 21.0%; Pred. No. 0.033;  
Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;

Qy 1 MKTKTLAVSASSLLMSANAIS-----YGSNADAP 34  
Db 15 MIENKXIVASALMAGSFQALADESKTGYTYGKAGASVMSIADORELSGNGEELSK 74  
Qy 35 YVAKIGQVADKQINGKNFAY--GIYAGYFNDQNGV---EAEFV--GSDAKFENAGVS 86  
Db 75 YKGG-----DGHDTVSGGIAAGYDYPQFSIPVTELEFYARGKADSKYNDKD 124  
Qy 87 PVKG-----DYKS-----FGAYGYTRYNFINTPTFAKRLGIAR-----TKDV-- 126  
Db 125 SMSGGYWRDLDKNEVSVNTLMNLAYDFRINDSAPTPWVSAG--IGYAKRHHOKTGTISTWD 183  
Qy 126 -----TSRNATTYSNKSQDTSLAGVGVEFKPLANVGVEASYNLSEDAANAIS 173  
Db 184 YGYSSGRESLSRSGSADNFAMSLGAGRYDYTPDIALDLSRILDADSSVS 236

RESULT 6  
A39185  
virulence protein pagC precursor - Salmonella typhimurium  
C:Species: Salmonella typhimurium  
C:Date: 08-Nov-1991 #sequence\_revision 08-Nov-1991 #text\_change 08-Oct-1999  
C:Accession: A39185  
R:Pulkkinen, W.S.; Miller, S.I.  
J. Bacteriol. 173, 86-93, 1991  
A:Title: A Salmonella typhimurium virulence protein is similar to a Yersinia enterocolit  
A:Reference number: A39185; MUID:91100323  
A:Accession: A39185  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-188 <PUL>  
A:Cross-references: GB:M55546; NID:g154232; PIDN:AAA27179.1; PID:g154233

Query Match 11.8%; Score 108; DB 2; Length 188;  
Best Local Similarity 24.2%; Pred. No. 0.03;  
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

Qy 1 MKTKTLAVSASSLLMSANAISYGSADAPYVGAKIGQVADKQINGKNATAGITAG 60  
Db 1 MKNTILSTLVITTSVLVNVAAQDTNAPSVCYARAKSKV--ODFKRTGYNVKT----- 54  
Qy 61 YNPDNGVEAEFYGSDAKFENAGVSPYKGDVKSFGA-----YGF----- 101

Db 54 -----RYEDDSPVSFISLSLYGDRQASGVVEPEGIHYHDKFEVYKSLMVG 102  
Qy 101 -YRYNFINTPTFAKRLGIAR-TKYDVTSRNATTYSNK--SDKTSLAGVGVEFKPLANV 156  
Db 103 AYRRS-DNFSLYALAGVGTAKATKEHSTODGDSFNSSNISKTKTFAAGACVQNNPLENI 161  
Qy 157 GVEASY---NYLSEDAANAISGAHLAF 180  
Db 162 VVDVGYEGSNISSTKINGFNWGVGYRF 188

RESULT 7  
MMBP2  
outer membrane porin lc precursor - phage PA2  
C:Species: phage PA2  
A:Note: host Escherichia coli  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 02-Jul-1998  
C:Accession: D25647  
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.  
J. Biol. Chem. 261, 12723-12732, 1986  
A:Title: Structure of the lc and mmpC outer membrane porin protein genes of lambdaoid  
A:Reference number: A25647; MUID:86304457  
A:Accession: D25647  
A:Molecule type: DNA  
A:Residues: 1-365 <BLA>  
C:Genetics:  
A:Gene: lc  
C:Superfamily: outer membrane protein phoE  
C:Keywords: membrane protein; porin; trimer  
F:1-23/domain: signal sequence #status predicted <SIG>  
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

Query Match 11.8%; Score 108; DB 1; Length 365;  
Best Local Similarity 26.6%; Pred. No. 0.064;  
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

Qy 1 MKTKTL-AYSASSLLMSANAISY-----GNSADAPYVAK 39  
Db 1 MKTLVAISAAVASLMSAQAAETYNKSLDLGKVNKAHYFSSNDADDGDTTYAR 60  
Qy 40 IGOVDKQINGKNATYGYAGYFNDQNGVEAEFVGSADAKFENAGVSPYKGDVKSFGAYG 99  
Db 61 LFKKETIINDLGTFFGWE-YEFKGN--RAESGSSKDKTRLAPAGIK-----FGDIG 111  
Qy 100 TYRYNFINTPTFAKGL-----GIATKYDV-----TSRN----- 130  
Db 112 SIDYGRNAGVAYDIGAMVDYLPEFGDPTWTCYDVMTCRTTGFAIRYNNDFGLVDGLNF 171  
Qy 130 ATTYSNKSQDKT---SLAGVGVEFKPLANVGVEASYNLSEDAANAISGAHLA 179  
Db 172 AAQYOGKNDKRDSPFDNYTEGNGDGF-----GRSATYEV-----EGFGIGATYA 213

RESULT 8  
MMBNC  
outer membrane porin mmpC precursor - Escherichia coli cryptic lambdaoid prophage DLp1  
C:Species: Escherichia coli  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
C:Accession: A25647; S66594; G64787  
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.  
J. Biol. Chem. 261, 12723-12732, 1986  
A:Title: Structure of the lc and mmpC outer membrane porin protein genes of lambdaoid  
A:Reference number: A25647; MUID:86304457  
A:Contents: mutant strain CS384  
A:Accession: A25647  
A:Molecule type: DNA  
A:Residues: 1-365 <BLA>  
R:Mandi, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.  
J. Mol. Biol. 257, 561-573, 1996  
A:Title: Holliday junction resolvases encoded by homologous rusa genes in Escherichia  
A:Reference number: S66579; MUID:96196428  
A:Accession: S66594





A56152

major 25k outer membrane protein precursor - Brucella abortus

C:Species: Brucella abortus

C:Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 08-Oct-1999

C:Accession: A56152

R:de Wergifosse, P.; Lintermans, P.; Limet, J.N.; Clockaert, A.

J. Bacteriol. 177, 1911-1914, 1995

A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton

A:Reference number: A56152; MUID:95204367

A:Accession: A56152

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-213 &lt;DEA&gt;

A:Cross-references: GB:X79284; NID:g769744; PIDN:CAA5872.1; PID:g769745

## Query Match

11.1%; Score 102.5; DB 2; Length 213;

Best Local Similarity 27.1%; Pred. No. 0.1;

Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;

QY 1 MKTKTLAVSASSLLAMANAISYGSADA---QPYVGAKI-----41

Db 1 MRLTKSLIVISA-ALPEPSATF-----ADAIOEPVPAPVVEAPQYSWAGTGLYL 54

QY 41 --GGVDAK-----QINGKNTAVGIYAGYFNPDN---FGVEAFVGSDAKEFNAGVSPVKG 90

Db 55 GYGNMKAKTSTVGSIKPDWKAGAPAGNFFQDDIIVYEGSDAGYSMAKSKSGLEVKOG 114

QY 91 DVKSGAVGYRYNFINTPFYAKGKLG-----AKTKYDV 125

Db 115 FEGSLGARVGYDLNPV-MPYLTLAGIAGSQIKLNGLDDESKFRVGTWAGAGLEAKLTQNI 173

QY 126 TSR---NATYTSNKSDDKTSIAG-----GVGVGFK 151

Db 174 LGREVEYRYTGKNN--YDLAGTVRNKLDIQDIRVGIYK 212

## RESULT 12

major outer membrane protein P2 - Haemophilus influenzae (isolate d2 and isolate 3dr)

C:Species: Haemophilus influenzae

A:Variety: isolate d2; isolate 3dr

C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Oct-1999

C:Accession: S68062; S68065; S43699

R:Duim, B.

submitted to the EMBL Data Library, June 1993

A:Reference number: S68062

A:Accession: S68062

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-386 &lt;DUIM&gt;

A:Cross-references: EMBL:X73376; NID:g475184; PIDN:CAA51796.1; PID:el192124; PID:g265428

A:Experimental source: isolate d2

A:Accession: S68065

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-386 &lt;DUIM&gt;

A:Cross-references: EMBL:X73379; NID:g475184; PIDN:CAA51796.1; PID:el192124; PID:g265428

A:Experimental source: isolate 3dr

R:Duim, B.; van Alphen, L.; Eljk, P.; Jansen, H.M.; Dankert, J.

Mol. Microbiol. 11, 1181-1189, 1994

A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membrane

A:Reference number: S43699; MUID:94293786

A:Accession: S43699

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 89-105;135-153;178-196;223-236, 'GD', 239-246;269-275, 'HY', 278-298 &lt;DUIM&gt;

A:Cross-references: EMBL:X73376

A:Experimental source: isolate d2

C:Keywords: membrane protein

## Query Match

11.1%; Score 101; DB 2; Length 386;

Best Local Similarity 24.1%; Pred. No. 0.26;

Matches 45; Conservative 28; Mismatches 74; Indels 40; Gaps 8;

QY 5 KTLAVSASSLLAMANAISYGSADAQPYVGAKIGV-----DAKING-----K 51

Db 3 KTLAALIVGAFASANAANAIVYNNEG-TNVELGRLSTITEGNSGTVDQEQHGLARNA 61

QY 52 NTAVGIYAGYFNPDNFGVE-----AEFVGSDAKEFNAGVSPVKGDVSKFAGTGYR 103

Db 62 GSRFHIAKTHNFGDGFYAGYLETRLVSDYPSSSDHFG-----GITTKYAY 108

QY 104 NFINTPFYAKGKLGIAKTKVD-VTSRNATTY---SNKSDKTSIAGVGVGFKPLANV-GV 158

Db 109 VTGKAKAFGEVVKLGAKKTIADGITSADKEYGVLNKKIILTNGMTVGYTYKGIGDGL 168

QY 159 EASYNYL 165

Db 169 VLGANYL 175

## RESULT 13

S68063

major outer membrane protein P2, antigenic variant d3 - Haemophilus influenzae

C:Species: Haemophilus influenzae

C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Oct-1999

C:Accession: S68063; S43700

R:Duim, B.

submitted to the EMBL Data Library, June 1993

A:Reference number: S68062

A:Accession: S68063

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-386 &lt;DUIM&gt;

A:Cross-references: EMBL:X73377; NID:g475180; PIDN:CAA51794.1; PID:el192122; PID:g265

A:Experimental source: isolate d3

R:Duim, B.; van Alphen, L.; Eljk, P.; Jansen, H.M.; Dankert, J.

Mol. Microbiol. 11, 1181-1189, 1994

A:Title: Antigenic drift of non-encapsulated Haemophilus influenzae major outer membr

A:Reference number: S43699; MUID:94293786

A:Accession: S43700

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 89-105;135-153;178-196;223-236, 'GD', 239-246;269-275, 'HY', 278-298 &lt;DUIM&gt;

A:Cross-references: EMBL:X73377

A:Experimental source: isolate d3

C:Keywords: membrane protein

Query Match 11.1%; Score 101; DB 2; Length 386;

Best Local Similarity 24.1%; Pred. No. 0.26;

Matches 45; Conservative 28; Mismatches 74; Indels 40; Gaps 8;

QY 5 KTLAVSASSLLAMANAISYGSADAQPYVGAKIGV-----DAKING-----K 51

Db 3 KTLAALIVGAFASANAANAIVYNNEG-TNVELGRLSTITEGNSGTVDQEQHGLARNA 61

QY 52 NTAVGIYAGYFNPDNFGVE-----AEFVGSDAKEFNAGVSPVKGDVSKFAGTGYR 103

Db 62 GSRFHIAKTHNFGDGFYAGYLETRLVSDYPSSSDHFG-----GITTKYAY 108

QY 104 NFINTPFYAKGKLGIAKTKVD-VTSRNATTY---SNKSDKTSIAGVGVGFKPLANV-GV 158

Db 109 VTGKAKAFGEVVKLGAKKTIADGITSADKEYGVLNKKIILTNGMTVGYTYKGIGDGL 168

QY 159 EASYNYL 165

Db 169 VLGANYL 175

## RESULT 14

S68064

major outer membrane protein P2, antigenic variant d4 - Haemophilus influenzae

C:Species: Haemophilus influenzae





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 20, 2000, 12:22:44 ; Search time 30.1 Seconds

(without alignments)  
182.123 Million cell updates/sec

Title: US-09-164-714-7

Perfect score: 912  
Sequence: 1 MKTLTLAVSASSLLAMSA.....SYNYLSEDAANAISLGAHLAF 180Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SWISSProt\_38:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	13.5	178	1	YE57_HAEIN
2	123	13.5	178	1	ATL_YEREN
3	115.5	12.7	182	1	ATL_YEREN
4	113	12.4	359	1	OM53_HAEIN
5	112.5	12.3	359	1	OM21_HAEIN
6	112.5	12.3	361	1	OM22_HAEIN
7	109	12.0	365	1	PORI_BPAP2
8	108	11.8	188	1	PAGC_SALTY
9	108	11.8	365	1	NMPC_ECOLI
10	105	11.5	240	1	OM31_BRUME
11	105	11.5	363	1	OM2B_HAEIN
12	103	11.3	371	1	OM25_HAEIN
13	102.5	11.2	213	1	OM25_BRUBA
14	102.5	11.2	1310	1	VAC3_HELPY
15	101	11.1	386	1	OM24_HAEIN
16	100.5	11.0	201	1	OM25_BRUVY
17	100	11.0	353	1	OM52_HAEIN
18	100	11.0	361	1	OM29_HAEIN
19	100	11.0	1287	1	VAC2_HELPY
20	100	11.0	1291	1	VAC4_HELPY
21	99.5	10.9	172	1	OMPX_ENTCL
22	99.5	10.9	367	1	OM2A_HAEIN
23	99	10.9	360	1	OM26_HAEIN
24	98.5	10.8	213	1	OM25_BRUNE
25	98.5	10.8	369	1	OM27_HAEIN
26	96	10.5	385	1	OM23_HAEIN
27	95.5	10.5	1290	1	VAC0_HELPY
28	95	10.4	211	1	ROPB_RHIVY
29	94.5	10.4	213	1	OM25_BRUCA
30	94.5	10.4	213	1	OM25_BRUME
31	94.5	10.4	213	1	OM25_BRUSU
32	90	9.9	1296	1	VAC1_HELPY
33	89.5	9.8	1585	1	YOBO_BACSU
34	89	9.8	171	1	OMPX_ECOLI

35	89	9.8	359	1	OMPA_SERMA	P04845 serratia ma
36	88	9.6	346	1	OMPA_ECOLI	P02934 escherichia
37	86	9.4	394	1	OMSL_SALTI	Q06110 salmonella
38	86	9.4	623	1	YF22_HAEIN	P44242 haemophilus
39	85.5	9.4	341	1	OMPU_VIBCH	P97085 vibrio chol
40	85.5	9.4	356	1	VATX_DICDI	P54641 dictyostel
41	84.5	9.3	351	1	OMPA_SHIDY	P02935 shigella dy
42	84.5	9.3	715	1	STC2_STANU	P17855 staphylococ
43	84	9.2	300	1	V43E_MYCPN	P75157 mycoplasma
44	84	9.2	1569	1	VPJA_ECOLI	P52143 escherichia
45	83.5	9.2	486	1	VAT1_ECOLI	P77199 escherichia

## ALIGNMENTS

```
RESULT 1
ID YE57_HAEIN STANDARD: PRT: 178 AA.
AC 057201: 005063;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN H11457 PRECURSOR.
GN H11457.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20.
RX MEDLINE: 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback F.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Georgiades N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- SIMILARITY: WEAK, TO N.GONORRHOEA OPACITY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U32823; AAC23104.1; -.
CC DR TIGR: H11457; -.
CC DR Hypothetical protein; signal.
CC KW Hypothetical protein; signal.
CC FT SIGNAL 1 19
CC FT CHAIN 20 178
CC FT SEQUENCE 178 AA; 19470 MW; 15C5C544FF49A9F CRC64;
CC -----
Query Match 13.5%; Score 123.5; DB 1; Length 178;
Best Local Similarity 27.5%; Pred. No. 0.00082;
Matches 52; Conservative 30; Mismatches 82; Indels 25; Gaps 8;
```

```
4 LKTLAVSASSLLAMSAANAISYGSADAQPYGAKIG--QVDAKQINKNKT---AYGTY 58
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1 MKRLTLVMTLFTLAL-----SAQAQWYVQDGLGSKSIDITHVNSNSPSFTQRLS 50
AGYNPDONGVEAGAEFGSDAKFEEN-AGVSPVKGDVKSFGAYCTGYRNFTNTEFYAKGKLG 117
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
59 VGIAPDKNFRILAVDYTTNGKVTANTADYDVSLKSGSLGTLTFYEDLADFKPYV--GVR 108
```



DT 01-NOV-1995 (Rel. 32, last annotation update)  
DE OUTER MEMBRANE PROTEIN P5 PRECURSOR (OMP P5) (FIMBRIN).  
GN OMPA OR OMP5.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
CC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-NTHI 1128;  
RX MEDLINE; 9422575.  
RA Strakova T., Kolatukudy P.E., Murwin D., Billy J., Leake E.,  
RA Iam D., Demaria T., Bakaletz L.;  
RT "Role of fimbriae expressed by nontypeable Haemophilus influenzae in  
RT pathogenesis of and protection against otitis media and relatedness  
RT of the fimbria subunit to outer membrane protein A.";  
RL Infect. Immun. 62:2002-2020(1994).  
CC -1- FUNCTION: ACTS AS A FIMBRIAL SUBUNIT.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; I08448; AAA24959.1; -.  
DR PRINTS; PRO1021; OMPADOMAIN.  
DR PROSITE; PS01068; OMPA; FALSE\_NEG.  
DR PRAM; PF00691; OmpA; 1.  
DR PRAM; PF01389; OmpA\_membrane; 1.  
KW Outer membrane; Transmembrane; Porin; Signal; Fimbria.  
FT CHAIN 1 21  
FT SIGNAL 1 21  
FT DISULEID 22 359 OUTER MEMBRANE PROTEIN P5.  
FT DOMAIN 278 322 OMPA-LIKE.  
FT SEQUENCE 359 AA; 36340 MW; 576B1C59B4818C37 CRC64;  
SQ  
  
Query Match 12.4%; Score 113; DB 1; Length 359;  
Best Local Similarity 28.2%; Pred. No. 0.013;  
Matches 55; Conservative 18; Mismatches 80; Indels 42; Gaps 11;  
  
QY 5 KTLAVASASSLSANANAISYGNADAPYVAKIGQ-----VDAKOING----- 51  
DB 3 KTAIALVAGLAASVAGAAPOENTF-----YAGVAKGGSGHDGINNGALIKKGLSSSNY 58  
QY 51 ---KNT-AYGIYAGYNF--DONGVEAEFVGS--AKENFAGVSPVKGADYKSFAGY--- 99  
DB 59 GYRNTFTYGVFGCYQLINODNFGIAELGIDYDGRAKLRGAKPKAKH--TNHGAYLSL 116  
QY 99 -GYRNFNINPFYAKGLGIATKTYD--VTSRNATYISKSDKTSLAGS---VGVEFK 151  
DB 117 KGYTE---VLDGLDYVAGAGVALVRSDYKEYEDANGTRDHKKHGRHTARASGLFVAGADYA 173  
QY 152 PLANVGEASYNYS 166  
DB 174 VLPFLAVRLLEYQWLT 188  
  
RESULT 5  
OM21\_HAEIN STANDARD; PRT; 359 AA.  
AC P43839;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).  
GN OMP2 OR HI0139.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
CC

CC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE; 95350630.  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -1- SIMILARITY: TO PROTOBACTERIUM STRAIN S89 OMPH.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32699; AAC21810.1; -.  
DR TIGR; HI0139; -.  
DR KW Outer membrane; Transmembrane; Porin; Signal.  
FT CHAIN 1 20  
FT SIGNAL 1 20 BY SIMILARITY.  
FT SEQUENCE 359 AA; 39375 MW; 7E9EFFC2BD50314 CRC64;  
SQ  
  
Query Match 12.3%; Score 112.5; DB 1; Length 359;  
Best Local Similarity 28.1%; Pred. No. 0.014;  
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;  
  
QY 5 KTLAVASASSLSANANAISYGNADAPYVAKIGOV-----DAKOING-----K 51  
DB 3 KTAIALVAGLAASVAGAAVYNNNG--TNVELGGLSLIAEDSNTVDNQKHAKLNQ 61  
QY 52 NTAVGIYAGYEDQNF-----GVAEFV-----GSDAKEFNAGVSPVKGADYKSFAGY 101  
DB 62 GSRFHAKTHTNNGDGFVYAGYLETFYTKASENGSD--NF-----GDITS-----KY 106  
QY 102 RYNFINTPFYAKGLGIATKTYD--VTSRNATYIS--NKSDKTSLAGS--VGVEFKPLANV 157  
DB 107 AYVTLGNKRAFGEVKGIRAKTIADGITSADKEYGVNLNSDYIPTSGNTVGYTFKGDGLV 166  
QY 158 VEASY 162  
DB 167 LGANT 171  
  
RESULT 6  
OM21\_HAEIN STANDARD; PRT; 361 AA.  
AC P20149;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).  
GN OMP2.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
CC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SERO TYPE B;  
RX MEDLINE; 89079316.  
CC

```

RA Munson R.S. Jr., Tolan R.W. Jr.:
RT "Molecular cloning, expression, and primary sequence of outer
RL membrane protein P2 of Haemophilus influenzae type b.";
RN Infect. Immun. 57:88-94(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 901358127.
RA Munson R.S. Jr., Bailey C., Grass S.;
RT "Diversity of the outer membrane protein P2 gene from major clones of
RL Haemophilus influenzae type b.";
RL Mol. Microbiol. 3:1797-1803(1989).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89173305.
RA Hansen E.J., Hasemann C., Clausell A., Capra J.D., Orth K.,
RA Moomaw C.R., Slaughter C.A., Lattner J.L., Miller E.E.;
RT "Primary structure of the porin protein of Haemophilus influenzae
RT type b determined by nucleotide sequence analysis.";
RL Infect. Immun. 57:1100-1107(1989).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03359; AAA24993.1; -.
DR EMBL; A09003; CA00819.1; -.
DR PIR; A30542; A30542.
DR PIR; S09622; S09622.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.
FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).
FT SEQUENCE 361 AA; 39701 MW; AF104C9CD8942D69 CRC64;
SQ
Query Match 12.3%; Score 112.5; DB 1; Length 361;
Best Local Similarity 28.1%; Pred. No. 0.014;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;
QY 5 KTLAVASASSLLMSANAISYGNADAPYVGAKEIGY-----DAKOING-----K 51
DB 3 KTLALIVGAFASANAAYVYNNEG-TNVELGGRSLIIAEOSNSTYVDNKKQHGALRNQ 61
QY 52 NTAAGTYAGYNFDNF-----GVAEAEV-----GSDAKFENAGVSPYKGVKSGAYGTY 101
DB 62 GSRFHFKATFNFCDFGYAAGYLETREVTAKASENGSD--NF-----GDITS-----KY 106
QY 102 RYVFINTPFAKCKLGIATKVD-VTSRNAVTS--NKSQKTLAAG-VGVGKPLANVG 157
DB 107 AYTTLNKAAGVEKLGRAKTIADGITSADKREXGVLNNSDIPTSGNTVGYTFKGDGLV 166
QY 158 VEASY 162
DB 167 LGANY 171
RESULT 7
PORT_BPPA2 STANDARD: PRT; 365 AA.
AC P07238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORIN PROTEIN LC PRECURSOR.
GN LC.
OS Bacteriophage PA-2.

```

```

OC viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86304457.
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and ompc outer membrane porin protein genes of
RL lambdaoid bacteriophage.";
RL J. Biol. Chem. 261:12723-12732(1986).
CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYSOGENIC STATE. IN E. COLI THE EXPRESSION OF OMPC AND OMPF
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02580; AAA23301.1; -.
DR PIR; D25647; MMBP2.
DR HSSP; P02931; IGFN.
DR PRINTS; PR00182; ECOLINEPORIN.
DR PRINTS; PR00183; ECOLIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
DR PFAM; PF00267; Gram-ve_porins; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.
FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).
FT SEQUENCE 365 AA; 40290 MW; 0FBC0531FB9C0205 CRC64;
SQ
Query Match 12.0%; Score 109; DB 1; Length 365;
Best Local Similarity 26.6%; Pred. No. 0.029;
Matches 62; Conservative 20; Mismatches 77; Indels 74; Gaps 11;
QY 1 MKTLTKILL-AVSASSLLMSANAISY-----GNSADAOPYVGAK 39
DB 1 MKRLTYAISVAASVLMAMSAQAAEITYNKDSNKLIDYGVNAKHFFSSNDADGDTTYAR 60
QY 40 IGVNDAKOINGKMTAVGIYGVNFDNFGVEAEFVGSDAKEFNAGVSPYKGVKSGAYG 99
DB 61 LGFKGEQINDQLTGFGOMP-YEFKGN--RAESQSSKDKTLARGLK-----FGDYG 111
QY 100 TYRNFINTPFAKCKL-----GIATKVDY-----TSRN----- 130
DB 112 SIDYGRNYGVAVDIGAMTDVLPFGGDTWTQTVDYFMTGRTTFATYRNDFGLVGLNF 171
QY 130 ATTYSNKSKDT---SLAGYGVGFKPLANYGVEASTNYLSEDAANAISLGAHLA 179
DB 172 AAQYQGNDRSDPDNYTEGNGDGF-----GFSATVEY-----EGFGIGATYA 213
RESULT 8
PAGC_SALTY STANDARD: PRT; 188 AA.
AC P23988;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VIRULENCE MEMBRANE PROTEIN PAGC PRECURSOR.
GN PAGC.

```



```

OS Salmomella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae:
OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14028;
RX MEDLINE: 91100323.
RA Pulkinen W.S., Miller S.I.;
RT "A Salmomella typhimurium virulence protein is similar to a Yersinia
RT enterocolitica invasion protein and a bacteriophage lambda outer
RT membrane protein."
RL J. Bacteriol. 173:86-93(1991).
CC -1- FUNCTION: ESSENTIAL FOR FULL VIRULENCE AND SURVIVAL WITHIN
CC MACROPHAGES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE ATL/OMPX/PAGC/LOM FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55546; AAA27179.1; -.
DR PIR: A39185; A39185.
DR STYGENE: SG77777; PAGC.
DR PRINTS: PR00316; ENTEROVIROMP.
DR PROSITE: PS00694; ENT_VIR_OMP_1; 1.
DR PROSITE: PS00695; ENT_VIR_OMP_2; 1.
KW Outer membrane; Transmembrane; Signal; Virulence.
FT SIGNAL 1 23
FT CHAIN 24 188
FT SEQUENCE 188 AA; 20574 MW; B23826FA2B62DDE CRC64;
SQ
Query Match 11.88; Score 108; DB 1; Length 188;
Best Local Similarity 24.28; Pred. No. 0.017;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;
QY 1 MKTLTLLAVASSILLASANAISYGSADAOPYVGAKIGIOYAKOINGRKTAAGIYAG 60
DB 1 MKNITLTLVTTLVAVNAADPNAFSGVARYAQSKV--QDEKINRGVAVK----- 54
QY 61 YNFDNFEVGEAFVGSDAKEFNAGVSPYKGVKSGA-----YGT----- 101
DB 54 -----RYEDDSVYSFISLSIYLXGDRQASGVPEPGIHHDKFEYKYGSLMVG 102
QY 101 -YRYNFTTPPYANGKIGIAF-TKVDVTSRNATYTSNK--SDKTSLAGGVGVKPLANV 156
DB 103 AYRLS-DNFSILYALAGVGTAKTEKEHSTODGDSFSKISSRKTFGAWAGVOMPLENI 161
QY 157 GVEASY---NYLSEDAANISGAHLAF 180
DB 162 VVDVGEYSNISTKINGENVGVGRF 188

```

```

RX MEDLINE: 86304457.
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;
RT "Structure of the lc and mmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage."
RL J. Biol. Chem. 261:12723-12732(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426517.
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federpspiel N., Hyman R., Kalman S., Komp C., Kirdi O., Lew H., Lin D.,
RA Nannath A., Oetner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-17 FROM N.A.
RC STRAIN-JL5502;
RX MEDLINE: 94335635.
RA Coll J.L., Heyde M., Portaller R.;
RT "Expression of the mmpc gene of Escherichia coli K-12 is modulated by
RT external pH. Identification of cis-acting regulatory sequences
RT involved in this regulation."
RL Mol. Microbiol. 12:83-93(1994).
RN [5]
RP SEQUENCE OF 347-365 FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 96196428.
RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
RT "Holiday junction resolvases encoded by homologous rusa genes in
RT Escherichia coli K-12 and phage 82."
RL J. Mol. Biol. 257:561-573(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- MISCELLANEOUS: IN WILD-TYPE STRAINS OF E. COLI K12, THE NMPC OPEN
CC READING FRAME IS INTERRUPTED BY AN IS5 INSERTION AND GENERATES A
CC HYBRID OPEN READING FRAME THAT IS NOT EXPRESSED. HOWEVER, IN
CC MUTANT STRAIN CS348, THE IS5 ELEMENT HAS BEEN DELETED AND NMPC IS
CC EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE OMPX/PHO FAMILY OF PORINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13457; AAA23728.1; ALT_SEQ.
DR EMBL: AE000160; AAC73654.1; ALT_SEQ.
DR EMBL: U82598; AAB40749.1; ALT_SEQ.
DR EMBL: Z35442; CAAB4594.1; -.
DR EMBL: X92587; CAAB3325.1; -.
DR PIR: A25647; MMECNC.
DR HSSP: P02931; IGFN.
DR ECOGENE: EG10659; NMPC.
DR PRINTS: PR00182; ECOLINPORIN.
DR PRINTS: PR00183; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
DR PFAM: PF00267; Gram-ve_porins; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365
FT CONFLICT 326 326 K -> N (IN REF. 2 AND 3).
FT SEQUENCE 365 AA; 40316 MW; 6E512BD4647B4F8 CRC64;

```

Query Match 11.8%; Score 108; DB 1; Length 365;  
Best Local Similarity 26.6%; Pred. No. 0.035;  
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;

QY 1 MKTLKTL-AVSASLLAMSANAISY-----GNSADAQPYVGAK 39  
DB 1 MKLTVAISNAVASVMAASQAELIYNDSKLDLXGVNAKHRYSSNDADGDTTYAR 60  
QY 40 ICQVDAKQINGKNTAYGIYVFNQDFVEAEFVSGDAKEFNAGVSPYKGDVKSFGAYG 99  
DB 61 LQFKGFTQINDLTGFGQME-YEFKGN---RAESQSSKDKTRLAPAGLK-----FGDYG 111  
QY 100 TYRYNININPFYAKGL-----GIATKYDV-----TSRN-----130  
DB 112 SLIDGRNYGVAYDIDIGAMTDVLPFGGDTWTQTDVMTORATGVATRYRNDPFGLDVGLNF 171  
QY 130 ATTYSKRSKPT---SLAGVGVGFKPLANVGVEASYNVISEDANATSLGHILA 179  
DB 172 MAQYQKNDRSDPDNTYEGNGDF-----GFSATLEY-----EGFGIGATYA 213

RESULT 10  
OM31\_BRUME STANDARD; PRT: 240 AA.  
AC Q45322;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE 31 KD OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.  
GN OMP31.  
OS Brucella melitensis.  
OC Brucella: Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-16M.  
RX MEDLINE: 96355886.  
RA Vizaino N., Cloeckaert A., Zygmunt M.S., Dubray G.;  
RT "Cloning, nucleotide sequence, and expression of the Brucella  
RT melitensis omp31 gene coding for an immunogenic major outer membrane  
RT protein.";  
RT Infect. Immun. 64:3744-3751(1996).  
CC -1- FUNCTION: MAJOR OUTER MEMBRANE PROTEIN ASSOCIATED WITH  
CC -1- PEPTIDOGLYCANS. MAY FUNCTION AS A PORIN.  
CC -1- SUBUNIT: OLIGOMERIC.  
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AF076290; AAB3693.1; -.  
DR PFAM: PF01389; OmpA membrane; 1.  
KW Antigen; Outer membrane; Porin; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 240 31 KD OUTER-MEMBRANE IMMUNOGENIC PROTEIN.  
FT DOMAIN 48 83 EPIOTOPE RECOGNIZED BY THE MONOCLONAL  
FT ANTI-BODY A59/10F09/G10.  
SQ SEQUENCE 240 AA; 25323 MW; 21C6E5BC479F66A5 CRC64;

Query Match 11.5%; Score 105; DB 1; Length 240;  
Best Local Similarity 24.4%; Pred. No. 0.04;  
Matches 51; Conservative 29; Mismatches 85; Indels 44; Gaps 9;

QY 4 LKTLAVSASLLAMSANA---ISYGNADAP-----YVAKTG-----42  
DB 1 MKSVLIASIAAMFATISAMADAVVSESPAPTAAPVDTFSWGTGYIGINAGYAGCKRHPF 60

QY 42 ----QVDAKQINCK-----NTAYGIYAGVNFQDNFV----EAEFVSGDAK-EFNAGVS 86  
DB 61 SRFDRKEDNEQVSGSLDVTAGGFGVQAGVNMQLDNGVYLGAETDPQGSVTSISAGAS 120  
QY 87 PYKGDVKS-FGAGTRYRYNINP-----FYAKGKIAGIAKTYDVT-SRRATTYSKSKDK 139  
DB 121 GLEKAEETVEMFGYVRAIRLGYATERLWYGTGGLAYKVSFAFNIGDASALHTWSDK 180  
QY 140 TSLAGVGVGFKPLANVGVEASYNVISED 168  
DB 181 TRAGMTLGAGAEIATINNMNTLASEVLYTD 209

RESULT 11  
OM2B\_HAEIN STANDARD; PRT: 363 AA.  
AC Q48220;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).  
GN OMP2.  
OS Haemophilus influenzae.  
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-3230B.  
RX MEDLINE: 94018553.  
RA Duim B., Dankert J., Jansen H.M., van Alphen L.;  
RT "Genetic analysis of the diversity in outer membrane protein P2 of  
RT non-encapsulated Haemophilus influenzae.";  
RT Microb. Pathog. 14:451-462(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: X73391; CAA51808.1; -.  
DR Outer membrane; Transmembrane; Porin; Signal.  
KW Antigen; Outer membrane; Transmembrane; Porin; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 363 OUTER MEMBRANE PROTEIN P2.  
SQ SEQUENCE 363 AA; 39873 MW; 2739FDBF42A3051 CRC64;

Query Match 11.5%; Score 105; DB 1; Length 363;  
Best Local Similarity 26.9%; Pred. No. 0.062;  
Matches 52; Conservative 29; Mismatches 78; Indels 34; Gaps 11;

QY 5 KTLAVSASLLAMSANAISYGNADAPYVGAKIGOV-----DAKQING-----K 51  
DB 3 KTLAALIVGAFASANAAMVAVVYNNEG-TNVELGGRSLIAEAGSNSTIKQKQHGALRMQ 61  
QY 52 NRYAYGIYAGVNFQDNF-----GVAEFVSGDAKEFNAGVSPYKGDVKSFGAYGT-YRYNPI 106  
DB 62 SSRFHAKATHNFGDGFYAGGYETRLVSAQ-----SGTESD--NFGHIIITRYKAVYVTL 111  
QY 107 NPFYVKGKIGIAKTYDVT-VTSRNATTYS--NKSOKTSLAGC-VGVGFKPLANVGVEASY 162  
DB 112 GNKAFGEVVKLGRAKTTADGITSADKEGYVLNNSKIPTINGNTGYTERGIDELVIGANT 171  
QY 163 NYLSEDAANAISLG 175  
DB 172 -LLAQERNKRYGTG 183

RESULT 12  
OM25\_HAEIN

```

ID OM25_HAEIN STANDARD; PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2.
RX MEDLINE: 94293786.
RA Duim B., van Alphen L., Eijk P., Jansen H.M., Dankert J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations."
RL MOL. Microbiol. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2.
RX MEDLINE: 94018553.
RA Duim B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae."
RL Microb. Pathog. 14:451-462(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73386; CAAS1803.1; -.
DR EMBL: X73383; CAAS1800.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT CHAIN 1 20
FT SIGNAL 21 371
FT VARIANT 184 184 K -> Q (IN T2).
FT VARIANT 224 224 A -> T (IN T2).
FT VARIANT 273 273 S -> T (IN T2).
SQ SEQUENCE 371 AA; 40901 MW; D0F252D7A6E5CDA6 CRC64;

Query Match 11.3%; Score 103; DB 1; Length 371;
Best Local Similarity 24.8%; Pred. No. 0.093;
Matches 53; Conservative 29; Mismatches 74; Indels 58; Gaps 12;

QY 5 KTLIAVSASSLLMSANAISYGSNADQPYVAKIGOV-----DAKOING--K 51
DB 3 KTLIAALIVGFAASAAVAAYVANNESG-SKVELGRLSVIAEQSNNTVDDQOQAGALRMQ 61
QY 52 NTAAYGIAGYNFDONF-----GVEAEFVG---SDAKFEFNAGVSPYKGDVSKFGAGTYRYN 104
DB 62 GSRFHIAKATNFNGGFGFYAGCYLETRFISHYQDNADHFD-----DITF-----KYATV 108
QY 105 FINTPFYAKKGLGIATKTV-DVTSRNATYVS--NKSOKTSLIAGV-VGVGKPLANVGVEA 160
DB 109 TLGKRAKAGEVYKLRKAKTIADITSAEDKEYGLVNSKYIRFTNGMTVGYTFEGID--GLVL 166
QY 161 SYNTLSEDA-----NAISLGA 176
DB 167 GANYTLAQAARDTANPGKKGVEAAGVISNGVQVGA 200

```

```

ID OM25_BRUBAB STANDARD; PRT; 213 AA.
AC Q44664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KD OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-544S;
RX MEDLINE: 95204367.
RA de Wergifosse P., Lintermans P., Linet J.N., Cloeckaert A.;
RT "Cloning and nucleotide sequence of the gene coding for the major 25-
RT kilodalton outer membrane protein of Brucella abortus."
RL J. Bacteriol. 177:1911-1914(1995).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79284; CAAS5872.1; -.
DR PFAM: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 213
SQ SEQUENCE 213 AA; 23052 MW; 2328515F1F794BC7 CRC64;

Query Match 11.2%; Score 102.5; DB 1; Length 213;
Best Local Similarity 27.1%; Pred. No. 0.057;
Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;

QY 1 MKTKTLIAVSASSLLMSANAISYGSNADA--QPYVAKI----- 41
DB 1 MKTKLSLYISA-ADLPESATAF-----ADAIDQBPVPAPVEYAPQYSNAGGTYGLYL 54
QY 41 --GOVDN-----QINGKNTAYGIAGYNFDON--FGVEAEFVSDAKFEFNAGVSPYK 90
DB 55 GYGNKRAKTSIVSGSIKPDMDKAGAFAGNPFQDDIYGVGEDAGYSMAKKSKDGLGVKQG 114
QY 91 DVKSGFAGVGYRYNFINTPFYAKKGLGI-----AKTKVDV 125
DB 115 FEGSLGKRVGYDLDPV-MPYLTLAGIAGQIKRLNGLNGLDESKFRVGTWAGLLEAKLTNDI 173
QY 126 TSR---NATYYSNKSOKTSLIAG-----GYVGK 151
DB 174 LGRVEYRYTQIGKNK--YDLAGTYVRNKLDTQDIRVIGIGYK 212

```

RESULT 13  
OM25\_BRUBAB

```

RX MEDLINE; 95355366.
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuu M.K., Blaser M.J.,
RT Cover T.L.;
RT "Mosaicism in vaccinating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vacA types with cytotoxin production and
RT peptic ulceration.";
RL J. Biol. Chem. 270:17771-17777(1995).
CC -1- FUNCTION: INDUCES VACCINATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U29401; AAA86834.1; -.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 2 VACUOLATING CYTOTOXIN.
FT PROPEP ? 1310 POTENTIAL.
SQ SEQUENCE 1310 AA; 14198 MW; 1BC21FE3D435F981 CRC64;

Query Match 11.2%; Score 102.5; DB 1; Length 1310;
Best Local Similarity 22.7%; Pred. No. 0.4; Mismatches 74; Indels 53; Gaps 10;
Matches 48; Conservative 36;

OY 9 AVSASSLLAMSAANAISYGNADAPYVGAKI-----GQVDAKQINGKN 52
DB 1049 AIGGTS-LNNGSNASL-YGTSAVDAYLNGEYEAIVGGRGYSGSFSQANSLSNGANN 1106
OY 53 TAYGIYAGNFDQN-FGVAAE-FVGSDAKEFNAGVSPVKGVKSF-----GAYGT 100
DB 1107 TNEGYSTRFAHQHEDEFAQALGSDOSLSLFKSLALDLQNSHYLYLAYSATTRASG- 1166
OY 101 YRYNTNPFYAKGLKIKTKYVDTSRNATYTSNKSDDKTSLAGVGVGFKPLANGVEA 160
DB 1166 YFAFRRNALVLRKPSGVSYNHL-----GSTNFKSNSNOVALSSSSOHLNANAVEA 1220
OY 161 SYNY-----LSEDA-----MAISL 174
DB 1221 RYYGDTSYFYMNAGVLQEFARFGSNNNAVSL 1251

RESULT 15
OM24_HAEIN STANDARD: PRT; 386 AA.
AC P46026;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D2, D3, D3R, AND D4;
RX MEDLINE; 94293786.
RA Duim B., van Alphen L., Eijk P., Jansen H.M., Dankert J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein p2 in patients with chronic bronchitis is
RT caused by point mutations.";
RL Mol. Microbiol. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D2, D3, D3R, AND D4;
RX MEDLINE; 94018553.

```

```

RA Duim B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein p2 of
RT non-encapsulated Haemophilus influenzae.";
RL Microb. Pathog. 14:451-462(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN D2/D3R IS SHOWN HERE.
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN S59 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X73376; CAAS1793.1; -.
DR EMBL; X73377; CAAS1794.1; -.
DR EMBL; X73378; CAAS1795.1; -.
DR EMBL; X73379; CAAS1796.1; -.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 386 OUTER MEMBRANE PROTEIN P2.
FT VARIANT 284 285 D -> G (IN D3).
FT VARIANT 284 285 DD -> GG (IN D4).
SQ SEQUENCE 386 AA; 41905 MW; 05F8909A2382A1F3 CRC64;

```

```

Query Match 11.1%; Score 101; DB 1; Length 386;
Best Local Similarity 24.1%; Pred. No. 0.14; Mismatches 74; Indels 40; Gaps 8;
Matches 45; Conservative 28;

OY 5 KTLAVASASSLLAMSAANAISYGNADAPYVGAKIGOV-----DAKQING-----K 51
DB 3 KTLALIVGAFAPASANAANAVYNNBEG-TNVELGGRSLITTEBSNSTVDDQDQHGALRNA 61
OY 52 NTAYGIYAGNFDQNGVE-----AEFVGSDAKEFNAGVSPVKGVKSGAGCTYRY 103
DB 62 GSRFHVKATNFGDGFYAGGLETRLVSDYPGSSSDHFC-----GITKKYAY 108
OY 104 NTINTNPFYAKGLKIKTKYVD-VTSRNATY---SKSKRTSLAGVGVGFPLANV-GV 158
DB 109 VTLGNKAFGEVRLGRAKTIDAGITSADREKGYVLNKKYILTNGMTVGYTKGIDLDGL 168
OY 159 EASYNYL 165
DB 169 VLGANYL 175

```

Search completed: May 20, 2000, 12:22:46  
Job time: 3960 sec



**THIS PAGE BLANK (USPTO)**

[illegible]





QY 169 ANAISGAHLAF 180  
Db 161 VGTWNGVGYRF 172

RESULT 5

085182 PRELIMINARY; PRT; 572 AA.

AC 085182; 01-NOV-1998 (TREMblrel. 08, Created)

DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)

DE FLAGELLIN A. 12, Last annotation update)

GN FLAA.

OS Campylobacter jejuni.

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;

OC Campylobacter.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-D2290;

RA MEINERSMANN R.J., HIETT K.L.;

RT "Concerted evolution of duplicate fla genes in Campylobacter.";

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF050188; AAC25641.1;

DR PFAM: PF00700; Flagellin\_C; 1.

DR PFAM: PF00669; Flagellin\_N; 1.

DR PRINTS: PR00207; FLAGELLIN.

SQ SEQUENCE 572 AA; 59021 MW; 3A8B8A00 CRC32;

Query Match 12.8%; Score 117; DB 2; Length 572;

Best Local Similarity 24.3%; Pred. No. 0.079;

Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AAISYGSADQAPVYVAKIGOVDAKQ-----INGKMTAVGIYA----- 60

Db 242 AAVAGATSDTFAINGVIGVYDKDGDANGALVAALNSVKTGTTGVEASIDANGQLLITS 301

QY 60 ----GYNFDONFGEAEFVGSDAKEFNAGVSPVKGDKV-----SFGAYGYRY 103

Db 302 REGGIRKIDNGIGGA-FINADMKENGRSLVKNDDGDLISSNLSAGFGA----- 355

QY 104 NFINTPYAKGLGIATK--VDYTSRNATTYSNKSDDKTSIAG-----GVGVGFKP 152

Db 355 ----TOFISQASVSLRESKQIDANIDAMGFGSANKGVVLGYSSVAYMSAGSGFSS 410

QY 153 LANYGVEASVNYLSEDAANAISLGA 176

Db 411 GSGYVSGSKNYSFGANALAIISA 434

RESULT 6

030689 PRELIMINARY; PRT; 572 AA.

AC 030689; 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DE FLAGELLIN A. 10, Last annotation update)

GN FLAA.

OS Campylobacter jejuni.

OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;

OC Campylobacter.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-D1118;

RA MEINERSMANN R.J., HIETT K.L.;

RT "Concerted evolution of duplicate fla genes in Campylobacter.";

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

DR PRINTS: PR00207; FLAGELLIN.

SQ SEQUENCE OF 101-189 FROM N.A.

RC STRAIN-D224;

RA MEINERSMANN R.J., HIETT K.L., FIELDS P.I., HIETT K.L.;  
RL J. Clin. Microbiol. 0:0-0(1997).  
DR EMBL: AF050197; AAC25650.1;  
DR EMBL: AF015092; AAB69354.1;  
SQ SEQUENCE 572 AA; 59049 MW; A624D1EE CRC32;

Query Match 12.8%; Score 117; DB 2; Length 572;

Best Local Similarity 24.3%; Pred. No. 0.079;

Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AAISYGSADQAPVYVAKIGOVDAKQ-----INGKMTAVGIYA----- 60

Db 242 AAVAGATSDTFAINGVIGVYDKDGDANGALVAALNSVKTGTTGVEASIDANGQLLITS 301

QY 60 ----GYNFDONFGEAEFVGSDAKEFNAGVSPVKGDKV-----SFGAYGYRY 103

Db 302 REGGIRKIDNGIGGA-FINADMKENGRSLVKNDDGDLISSNLSAGFGA----- 355

QY 104 NFINTPYAKGLGIATK--VDYTSRNATTYSNKSDDKTSIAG-----GVGVGFKP 152

Db 355 ----TOFISQASVSLRESKQIDANIDAMGFGSANKGVVLGYSSVAYMSAGSGFSS 410

QY 153 LANYGVEASVNYLSEDAANAISLGA 176

Db 411 GSGYVSGSKNYSFGANALAIISA 434

RESULT 7

051841 PRELIMINARY; PRT; 346 AA.

AC 051841; 01-JUN-1998 (TREMblrel. 06, Created)

DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)

DE OUTER MEMBRANE PROTEIN 34 PRECURSOR.

GN OMP34.

OS Actinobacillus actinomycetemcomitans

OC (Haemophilus actinomycetemcomitans).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Actinobacillus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NCTC 9710;

RX MEDLINE: 98084499.

RA WHITE P.A., NAIR S.P., KIM M.J., WILSON M., HENDERSON B.;

RT "Molecular characterization of an outer membrane protein of

RT Actinobacillus actinomycetemcomitans belonging to the OmpA family.";

RL Infect. Immun. 66:369-372(1998).

DR EMBL: AF005079; AAC00068.1;

DR HSSP: P02934; 18XW.

DR PROSITE: PS01068; OMPA; 1.

DR PFAM: PF00691; OmpA; 1.

DR PFAM: PF01389; OmpA-membrane; 1.

DR PRINTS: PR01021; OMPADOMAIN.

KW SIGNAL.

FT SIGNAL. 1 21 POTENTIAL.

SQ SEQUENCE 346 AA; 36905 MW; 8A80E0A3 CRC32;

Query Match 12.7%; Score 116; DB 2; Length 346;

Best Local Similarity 25.3%; Pred. No. 0.05;

Matches 49; Conservative 22; Mismatches 75; Indels 48; Gaps 9;

QY 5 KTLAVSASSILAMSANAAISYGSADQAPVYVAKIGOVDA-----KQ-----INGKNT 53

Db 3 RTAIALATAGIAAATVAQAPDANTF-----YAGAKAGASSHHGLNDFKQGVSTINRSE 58

QY 54 AYGIYAGYNFQNGVEA--EFVGSDAKEFNAGVSPVKGDKVSGAYGYRYNFINTPFY 111

Db 59 AYGVFGYQIDNFAVEAGYEFGRSKAKVN-----GAQ-RFRHTAIGTTLA 104

QY 112 AK-----GKIGIAKTKVDYTSRNAT--TYSNKSDDKTSIAGGVGVGFKPLANV 156

DB 105 LKASPYLDNLVYGVGALLRSDYKVGQANKPDRYHMLKVPVAGGEVAILP--EL 162  
 QY 157 GVEASNYLSEDA 170  
 DB 163 ALRAEYQWVSRYGN 176

## RESULT 8

Q46679 PRELIMINARY: PRT: 250 AA.

AC 046679: 046679: PRELIMINARY: PRT: 250 AA.  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE HEAT RESISTANT AGGLUTININ 1.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=O9:H10:K99;  
 RX MEDLINE: 95012721.  
 RA LUTWYCH P., RUPPS R., CAVANAGH J., WARREN R.A., BROOKS D.E.;  
 RT "Cloning, sequencing, and viscometric adhesion analysis of heat-  
 RT resistant agglutinin 1, an integral membrane hemagglutinin from  
 RT Escherichia coli O9:H10:K99."  
 RL Infect. Immun. 62:5020-5026(1994).  
 DR EMBL: U07174; AAC13752.1; -;  
 DR PFAM: PF01389; OmpA\_membrane; 1.  
 DR CHAIN 26  
 FT SEQUENCE 250 AA; 27477 MW; 9673F116 CRC32;  
 SQ

Query Match 12.0%; Score 109.5; DB 2; Length 250;

Best Local Similarity 21.0%; Pred. No. 0.11; Mismatches 75; Indels 71; Gaps 10;

Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;  
 QY 1 MKTLKTLAVSASSLAMSAANAIS-----XGNSDAQP 34  
 DB 1 MEMKNVIAVSLAMAGMSTQALADESKTGYVYTGKAGASVMSLADQRFSLNGEETSK 60  
 QY 35 YVAKTIGVDAKOINGNKPAY--GIYAGYNPDONGV---EAEPV--GSDAKEFNAGVS 86  
 DB 61 YKGG-----DGHDFVFGSGIAGYDFYPOFSIPVTELEFYARGLKADSKYVNDKD 110  
 QY 87 PVKG-----DVKS-----FGVGTFRYNFNTPEFYAKGLGIK-----TKVDV-- 126  
 DB 111 SMSGGYWRDLDLNEVSVNTLMNLNAYDFRNDSAFTYVWSAG--IGAKELHOKTTGISTWD 169  
 QY 126 -----TSRNATYYSNKSDKTSLAGVGFGFKPLANVGVEASNYLSEDAANAIS 173  
 DB 170 YGSGSGRSLSSRSSGADNFAMSLGAGVRYDVTPIALDLSTRYLDAGSSVS 222

## RESULT 9

Q46678 PRELIMINARY: PRT: 264 AA.

AC 046678: 046678: PRELIMINARY: PRT: 264 AA.  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE HEAT RESISTANT AGGLUTININ 1 PRECURSOR.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=O9:H10:K99;  
 RX MEDLINE: 95012721.  
 RA LUTWYCH P., RUPPS R., CAVANAGH J., WARREN R.A., BROOKS D.E.;  
 RT "Cloning, sequencing, and viscometric adhesion analysis of heat-  
 RT resistant agglutinin 1, an integral membrane hemagglutinin from

RT Escherichia coli O9:H10:K99."  
 RL Infect. Immun. 62:5020-5026(1994).  
 DR EMBL: U07174; AAC13751.1; -;  
 DR PFAM: PF01389; OmpA\_membrane; 1.  
 KR Signal.  
 FT SIGNAL 1 86  
 FT CHAIN 40 264  
 FT SEQUENCE 264 AA; 29024 MW; 77A084A2 CRC32;  
 SQ

Query Match 12.0%; Score 109.5; DB 2; Length 264;

Best Local Similarity 21.0%; Pred. No. 0.12; Mismatches 75; Indels 71; Gaps 10;

Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;  
 QY 1 MKTLKTLAVSASSLAMSAANAIS-----XGNSDAQP 34  
 DB 15 MEMKNVIAVSLAMAGMSTQALADESKTGYVYTGKAGASVMSLADQRFSLNGEETSK 74  
 QY 35 YVAKTIGVDAKOINGNKPAY--GIYAGYNPDONGV---EAEPV--GSDAKEFNAGVS 86  
 DB 75 YKGG-----DGHDFVFGSGIAGYDFYPOFSIPVTELEFYARGLKADSKYVNDKD 124  
 QY 87 PVKG-----DVKS-----FGVGTFRYNFNTPEFYAKGLGIK-----TKVDV-- 126  
 DB 125 SMSGGYWRDLDLNEVSVNTLMNLNAYDFRNDSAFTYVWSAG--IGAKELHOKTTGISTWD 183  
 QY 126 -----TSRNATYYSNKSDKTSLAGVGFGFKPLANVGVEASNYLSEDAANAIS 173  
 DB 184 YGSGSGRSLSSRSSGADNFAMSLGAGVRYDVTPIALDLSTRYLDAGSSVS 226

## RESULT 10

Q33796 PRELIMINARY: PRT: 165 AA.

AC 033796: 033796: PRELIMINARY: PRT: 165 AA.  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE ATTACHMENT AND INVASION PROTEIN HOMOLOG.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RA DE GROOTE M., OCHSNER U.A., SHILOH M., NATHAN C., MCCORD J.M.,  
 RA DINADER M.C., LIBBY S.J., VAZQUEZ-TORRES A., XU Y., FANG F.C.;  
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBSJ databases.  
 DR EMBL: AF007380; AAB62386.1; -;  
 DR SEQUENCE 165 AA; 17439 MW; C70D1CB3 CRC32;  
 SQ

Query Match 12.0%; Score 109; DB 2; Length 165;

Best Local Similarity 25.6%; Pred. No. 0.07; Mismatches 68; Indels 32; Gaps 6;

Matches 41; Conservative 19; Mismatches 68; Indels 32; Gaps 6;  
 QY 36 YGAKTIGVDAKOING-KNT-----AYGIYAGYNF---DONGVEAEFV 74  
 DB 10 VGIAGISIGVRNAGYKKNVYSIGYATDLSGMLSGANANIKYNNEDDLSGGMAGSVT 69  
 QY 75 GSDAKEFNAGVSPKGVKVSFGAYGYRRN-FINPEFYAKGLGIKAKTVVDTSRNATY 133  
 DB 70 YNSADVNNNGYKKGDADYSLVGPSPYRPNDYLN---AYVMIGANGHI-----KDNW 119  
 QY 134 SNKSDKTSLAGVGFGFKPLANVGVEASNYLSEDAANAIS 173  
 DB 120 GNSDNKTAFAFGAGIQLNPVENIIVANVASYEHTSFTDADS 159

## RESULT 11

Q46680 PRELIMINARY: PRT: 247 AA.

AC 046680: 046680: PRELIMINARY: PRT: 247 AA.  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

```
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HEAT RESISTANT AGGLUTININ 1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=09:H10:K99;
RX MEDLINE; 95012721.
RA LOTWYCHE P., RUPPS R., CAVANAGH J., WARREN R.A., BROOKS D.E.;
RT "Cloning, sequencing, and viscometric adhesion analysis of heat-
RT resistant agglutinin 1, an integral membrane hemagglutinin from
RT Escherichia coli O9:H10:K99."
RL Infect. Immun. 62:5020-5026(1994).
DR EMBL; U07174; AAC13753.1; -.
DR PFAM; PF01389; OmpA-membrane; 1.
FT CHAIN 23 247 HEAT RESISTANT AGGLUTININ 1.
SQ SEQUENCE 247 AA; 27104 MW; 580331BA CRC32;

Query Match 11.9%; Score 108.5; DB 2; Length 247;
Best Local Similarity 20.9%; Pred. No. 0.13;
Matches 48; Conservative 38; Mismatches 73; Indels 71; Gaps 10;

OY 4 KTLIAVSASSLLAMANAIS-----YGSADAQPYVG 37
DB 1 MNKYIVASALMAGNFSTQALDESKGTGYTGKAGASVMSIADQRFSLGGEETSKYKG 60
OY 38 AKIGOVDAKQINGKNTAY--GIYAGYNFDQNGV---EAEFV--GSDAKFNAGVSPVK 89
DB 61 G-----DGHDTYVSGGIIAGYDIPYPOFSIIVRTLEFYARGKADSKYNVDKDSMS 110
OY 90 G-----DVKS-----FGAYGYRYNFINTPFYAKGLGIK-----TKVDV----- 126
DB 111 GGYMRDLDKNEVSVNTLMLNAYDFRNDSAFTPWVSAG--IGYAKETIHKKTGISITWDGY 169
OY 126 --TSRNATTYSNKSDKTSLAGVGVGFKPLANVGEASYNLSEDAANAIS 173
DB 170 GSSGRESLSRSGSADNFAMSLGAGVRYDPTDIALDLSTRYLDAGDSSVS 219

RESULT 12
OY 048225 PRELIMINARY; PRT; 357 AA.
AC 048225;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P2 (FRAGMENT).
GN P2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3221B.
RX MEDLINE; 94018553.
RA DUTM B., DANKERT J., JANSSEN H., VAN ALPHEN L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae."
RL Microb. Pathog. 14:451-462(1993).
DR EMBL; X73392; CAA51809.1; -.
FT NON_TER 357
SQ SEQUENCE 357 AA; 39093 MW; 95AD9567 CRC32;

Query Match 11.5%; Score 104.5; DB 2; Length 357;
Best Local Similarity 24.8%; Pred. No. 0.43;
Matches 52; Conservative 28; Mismatches 77; Indels 53; Gaps 10;

OY 5 KTLIAVSASSLLAMANAISYGSADAQPYVGAKIGV-----DAKQING-----K 51
||||| : | :||| : | : : : : | : | : |
```

```
DB 3 KTLIAIVGAFASANAANAAYVNNEG--TKVELGRVSIIEOSTSGDDQKHQHALRNQ 61
OY 52 NTATGITYAGYNFDQNF-----GYEAEFVGSDAKEFNAGVSPVKGYKSGATGT--RYNFI 106
DB 62 GSRFHAKATHNCGDFYAGQYLETRFV--TDAS-----KNGSDNCGDITTKYAYFTL 111
OY 107 NMPFAKGLKIATKRVYTSNRNATTYSNK-----SDKTSIAGVGVGFKRLANVGVA 160
DB 112 GNKAFGEVRLGAKTIAD---GTTAEKDQGYLDSKKYIPTNGVNTIGITYKVTDLVL 167
OY 161 SYNYL-----SEDANAISLGA 176
DB 168 GANYLLAQRNVDGEVLPQKVSNGVYGA 197

RESULT 13
OY 046760 PRELIMINARY; PRT; 251 AA.
AC 046760;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE T1A INVASION DETERMINANT PRECURSOR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H10407;
RA FLICKENSTEIN J.M., KOPECKO D.J., WARREN R.L., ELSTINGHORST E.A.;
RL Infect. Immun. 64:0-0(1997).
DR EMBL; U20318; AAB0592.1; -.
DR PFAM; PF01389; OmpA-membrane; 1.
KW signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 251 T1A INVASION DETERMINANT.
SQ SEQUENCE 251 AA; 27992 MW; 5D0E54EC CRC32;

Query Match 11.4%; Score 104; DB 2; Length 251;
Best Local Similarity 22.6%; Pred. No. 0.3;
Matches 51; Conservative 40; Mismatches 79; Indels 56; Gaps 10;

OY 1 KTKTLIAVSASSLLAMANAISYGSADAQPYVGAKIG-----QVDA----- 46
DB 1 MEMKKVIVASALMAGMFSQAOLA--DESKTGFYVTGKAGASVYMQDQRFROFGDDV 58
OY 46 ---KQINGKNTAY--GIYAGYNFDQNGV---EAEFVGSDAKEFNAGVSPVKGYKSG 96
DB 59 YKKKGDKNDIVFGAGLAVGYDFYGHYVNPVTEVEEYFGRGADSRYYTLDTWRSPPGDGG 118
OY 97 AYGYR-----YNFIN---TPFYAKGLGIK-----TKVD-----VTS 127
DB 119 REDTQNRSLSVNTLMNTYYDFRNSSAFTPWVSAG--LGYARVHHKATYIDTSMNSESGLSD 177
OY 128 RNATTYSNKSDKTSLAGVGVGFKPLANVGEASYNLSEDAANAIS 173
DB 178 ISALHSGYDNMFAMSIGAGVRYDPTDIALDLSTRYLDAGKSSLS 223

RESULT 14
OY 092H01 PRELIMINARY; PRT; 1291 AA.
AC 092H01;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE VACUOLATING CYTOTOXIN PRECURSOR.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN-F38.  
 RX MEDLINE: 98453456.  
 RA ITO Y., AZUMA T., ITO S., SUTO H., MIYAJI H., YAMAZAKI Y., KOHLI Y.,  
 RA KURIYAMA M.;  
 RT "Full-length sequence analysis of the vacA gene from cytotoxic and  
 noncytotoxic Helicobacter pylori."  
 RL J. Infect. Dis. 178:1391-1398(1998).  
 DR EMBL: AF049643; AAD04279.1; -  
 SQ SEQUENCE 1291 AA; 139427 MW; 9E3761BB CRC32;

Query Match 11.4%; Score 104; DB 2; Length 1291;

Best Local Similarity 25.4%; Pred. No. 2.5;  
 Matches 54; Conservative 28; Mismatches 75; Indels 56; Gaps 12;

OY 9 AVSASSILAMSAANAISYGSADAPYVGAKI-----GQVDAKQINGKNTAYGIYAGY 61  
 Db 1029 AIGGASLNS-GGNTSL-YGTSAGVDALNKEVLAIVGGFGSYGSSFN--NQANSLNSGA 1084  
 OY 62 NFDQNGV-----EAEF-----VSDAKFENAGVSPVKGVKSGF--AYGT----- 101  
 Db 1085 N-NANFGVYSRIIPANRHEFDEFOAGAVGSDQSLNFKSALLRDLNOSYNVLAAGATRAS 1143  
 OY 101 --YRYNFIPTFAKGLGIATKVDVTSRNATYNSKSDKISLAGVGVGFKPLANVGY 158  
 Db 1144 YGYDFAFFRNALVLKPSVGSYNHLSGTNFE--SNSTNKTALKNGASSOHLFNASANY 1199  
 OY 159 EASYNY-----LSE-----DANAISL 174  
 Db 1200 EARYYYGDTSYFYMNAGVLOEFANFGSSNAVSL 1232

## RESULT 15

ID 09ZHT8 PRELIMINARY; PRT: 1291 AA.  
 AC 09ZHT8;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, last annotation update)  
 DE VACUOLATING CYTOTOXIN PRECURSOR.  
 GN VACA.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-F64.  
 RX MEDLINE: 98453456.  
 RA ITO Y., AZUMA T., ITO S., SUTO H., MIYAJI H., YAMAZAKI Y., KOHLI Y.,  
 RA KURIYAMA M.;  
 RT "Full-length sequence analysis of the vacA gene from cytotoxic and  
 noncytotoxic Helicobacter pylori."  
 RL J. Infect. Dis. 178:1391-1398(1998).  
 DR EMBL: AF049647; AAD04282.1; -  
 SQ SEQUENCE 1291 AA; 139395 MW; 2B366B27 CRC32;

Query Match 11.4%; Score 104; DB 2; Length 1291;

Best Local Similarity 25.4%; Pred. No. 2.5;  
 Matches 54; Conservative 28; Mismatches 75; Indels 56; Gaps 12;

OY 9 AVSASSILAMSAANAISYGSADAPYVGAKI-----GQVDAKQINGKNTAYGIYAGY 61  
 Db 1029 AIGGASLNS-GGNTSL-YGTSAGVDALNKEVLAIVGGFGSYGSSFN--NQANSLNSGA 1084  
 OY 62 NFDQNGV-----EAEF-----VSDAKFENAGVSPVKGVKSGF--AYGT----- 101  
 Db 1085 N-NANFGVYSRIIPANRHEFDEFOAGAVGSDQSLNFKSALLRDLNOSYNVLAAGATRAS 1143  
 OY 101 --YRYNFIPTFAKGLGIATKVDVTSRNATYNSKSDKISLAGVGVGFKPLANVGY 158  
 Db 1144 YGYDFAFFRNALVLKPSVGSYNHLSGTNFE--SNSTNKTALKNGASSOHLFNASANY 1199

OY 159 EASYNY-----LSE-----DANAISL 174  
 Db 1200 EARYYYGDTSYFYMNAGVLOEFANFGSSNAVSL 1232

Search completed: May 20, 2000, 12:22:04  
 Job time: 4021 sec



**THIS PAGE BLANK (USPTO)**